

AD-A089 253

CONTINENTAL ARMY COMMAND, FORT MONROE VA
CONARC TRAINING WORKSHOP, FORT GORDON, GEORGIA 5-7 OCTOBER 1971--ETC(U)
OCT 71 R E HAINES, I A HUNT, W C OSBORN

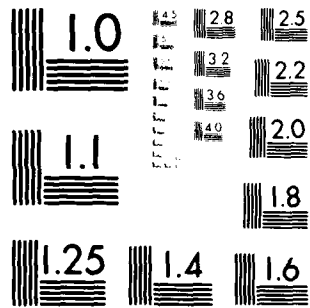
F/6 5/9

UNCLASSIFIED

NL

1 of 1
AD-A089 253

END
DATE
FILMED
10-80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A089253

CONFIDENTIAL

6 CONARC TRAINING WORKSHOP

FORT GORDON, GEORGIA,

5-7 OCTOBER 1971.

Sponsored by

US Continental Army Command

Hosted by

US Army Southeastern Signal School

9 Final Report In Seven Volumes

CONARC-CAL
October 1971
VOLUME VI.

Quality Control Speciality Workshop

10 GENERAL RALPH E. HAINES, JR.

Commanding General

US Continental Army Command

SEP 18 1980

A

BRIGADIER GENERAL IRA A. HUNT, JR.

Deputy Chief of Staff for Individual Training

US Continental Army Command

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

DDC FILE COPY

CONARC TRAINING WORKSHOP QUALITY CONTROL

CONARC TRAINING WORKSHOP
Fort Gordon, Georgia
5-7 October 1971

Volume IV - Quality Control Speciality Workshop

CONTENTS

	Page
1. Workshop Committee	VI-1
2. Speciality Workshop Summary	VI-2
3. Speciality Workshop Schedule	VI-5
4. Biographic Sketch - Dr. Alfred S. Drew	VI-7
5. Developing Performance Test for Training Evaluation	VI-8
6. Design of Objectively Scored Performance Tests and Use of Test Feedback	VI-15
7. Problems in Design of Subjectively Scored Performance Tests	VI-25
8. Test Scoring/Grading	VI-29
9. Criterion-Referenced Testing	VI-35
10. Go/No-Go Grading	VI-40
11. Promotion of Enlisted Personnel and Designation of Honor Graduates	VI-46
12. External Quality Control	VI-48

Accession For	
DATE	
BY	
REMARKS	
ART	
A	

CONARC TRAINING WORKSHOP .
QUALITY CONTROL

WORKSHOP COMMITTEE

Major William V. Green	US Army Adjutant General School
Mr. William C. Osborn	HumRRO Division 2
Mr. Robert N. Johnson	US Army Adjutant General School
Mr. James A. Squires	US Army Air Defense School
Mr. James L. Sherrill	US Army Adjutant General School
Mr. Earl Carr	US Army South- eastern Signal School
Dr. Alfred A. Drew	Purdue University Lafayette, Indiana

EXECUTIVE SUMMARY
FOR
Specialty Workshop: Quality Control

1. Abstract of presentation:

a. Discussion of a general model for the design of criterion tests (Mr. W. Osborn, HumRRO, 30 Min).

*b. Illustration of criterion test design for purposes of providing specific feedback to instructors and to training managers. Need for, and applications of ADP support. (Mr. R. Johnson, USAAGS, 45 Min).

*c. Issues in evaluation of subjectives skills (oral briefings, written communications, etc.). Problems in reliability and obtaining meaningful feedback information (Mr. R. Johnson, USAAGS, 20 Min).

*d. Issues in scoring and grading, weighting test items, and establishing standards. Considerations in remedial training, recycling, and failure. (MAJ W. Green and Mr. J. Sherrill, USAAGS, 30 Min).

e. Origins of norm referenced evaluation. Advantages of criterion referenced evaluation (Mr. Squires, USAADS, 30 Min).

*f. A method of obtaining grades based on completion times in self-paced courses (Mr. E. Carr, USASESS, 30 Min).

*g. Related issues in training quality control; overtraining, use of other feedback sources: student comments, postgraduate questionnaires, MODB reports, classroom observations. (MAJ W. Green, USAAGS).

h. Concluding comments and observations of the Quality Control Workshop (Dr. A. Drew, Purdue University).

NOTE: Above is the sequence of the precedings. Discussion periods followed presentations marked with an asterik.

2. Summary of results:

a. On the application of criterion referenced evaluation and training quality control.

✓
(1) There was almost complete agreement that Army service schools should adopt criterion referenced evaluation as the means of insuring that the trainee can do each job for which he is trained.

✓ *also in*
(2) ~~There was~~ agreement that superior performance on one task could not compensate for inadequate performance on another and that in effect our courses should be "multiple hurdle". There was recognition that this goal presented problems in group paced courses, because time constraints cause problems in handling failures on an individual basis and because it is difficult to obtain recognition (on manpower surveys) of staffing to support remedial training.

(3) No one disagreed with the need for ADP support of training quality control. Those schools without ADP support suggested they could not conduct an efficient and/or effective quality control program without ADP support.

(4) There was agreement that detailed feedback to the instructional system was of major value.

b. There was no disagreement that standards (absolute, fixed) are required. Considerable concern was expressed over the determination of such standards. Opinions ranged all the way from "Should be journeyman" to "Should be entry or apprentice," or "100% proficiency-no errors!" to "quite a few errors are allowable."

✓
c. The workshop attendees believed the present requirement to identify the honor graduate, the honor students and the upper half of each class as eligible for promotion produces inequities. Whether real or imagined, this proved to be an issue of considerable concern. There were three main aspects to this issue. ✓

(1) Failures are often promoted (by the company and the commander in the field) before successful graduates.

(2) Whether a student is designated eligible for promotion may depend on the class in which he is enrolled (one class may have a lower average than another).

(3) These requirements are not compatible with criterion referenced evaluation. Opinion here ranged from; "All such requirements should be eliminated" to "Even though awkward, sufficient information can be obtained to satisfy the requirement."

d. The Southeastern Signal School approach to identifying "honor students" in their self-paced course (based on completion times) was recognized as an excellent approach.

e. Evaluation of subjective tasks (briefings, staff studies, etc.) was recognized as highly subjective and while many solutions were offered, none were agreed to.

f. The recurring problem of postgraduate questionnaire return rates (10% to 80%) was noted.

g. Terminology problems again surfaced.

CONARC TRAINING WORKSHOP
5-7 October 1971

US Army Southeastern Signal School (USASESS)
Fort Gordon, Georgia

5 October General Session

0900-0910 Welcome by Post Commander
0911-0920 Welcome by Commandant
0921-0950 Mission and Purpose (CG CONARC)
0951-1030 Keynote Speaker
1031-1100 Break
1101-1110 Administrative Announcements
1111-1215 Conference Orientation by Specialty Area
 (Quality Control - MAJ N. Wood, 10 Min)
1216-1359 Lunch

QUALITY CONTROL
Workshop Session #1
"Training Accountability"

5 October 6 October

1400-1410	1400-1410	Workshop Introduction - MAJ N. Wood
1411-1440	1411-1440	HumRRO Presentation - Performance Test Development - Mr. W. Osborn
1441-1530	1441-1530	Design of Objectively Scored Performance Tests and Use of Feedback - Mr. R. Johnson
1531-1550	1531-1550	Break

Workshop Session #2
"Diagnosing Training Problems"

1551-1625	1551-1625	Workshop Discussion - Mr. W. Osborn/Mr. R. Johnson
1626-1635	1626-1635	Problem Areas in Designing Subjectively Scored Tests - Mr. R. Johnson
1636-1700	1636-1700	Workshop Discussion - Mr. R. Johnson

Workshop Session #3
"Criterion Testing"

<u>6 October</u>	<u>7 October</u>	
0830-0900	0830-0900	Test Scoring/Grading - Mr. J. Sherrill/ MAJ W. Green
0901-0930	0901-0930	Discussion
0931-1000	0931-1000	Air Defense School Presentation on Criterion Testing - Mr. Squires
1001-1030	1001-1030	S. E. Signal School Presentation on GO/NO GO Grading - Mr. E. Carr
1031-1100	1031-1045	Break

Workshop Session #4
"Problem Areas and External Quality Control Sources"

1101-1115	1046-1100	Enlisted Promotion and Honor Graduate Problem Areas - MAJ W. Green
1116-1130	1101-1115	Other Types of Quality Control - MAJ W. Green
1131-1150	1116-1135	Workshop Discussion
1151-1215	1136-1200	Workshop Critique and Observations - Dr. A. Drew
1216-1400	1201-1330	Lunch
	1331-1445	Summaries by Specialty Area (Quality Control, Mr. J. Sherrill)
	1446-1500	Critique
	1501-----	Depart

CONARC Training Workshop Quality Control

Civilian Consultant - Dr. Alfred S. Drew

Dr. Drew is currently Professor of Industrial Education at Purdue University. He teaches and counsels at the graduate and undergraduate levels and helps direct a special training program for labor. He served as Chairman of Vocational-Technical Education, as well as Chairman of the Graduate Education Curriculum Committees. He was Project Director of Purdue University's National Apprenticeship Study sponsored by the US Department of Labor.

Dr. Drew received his PHD from the University of Wisconsin, majoring in Adult Education - Curriculum Planning. His dissertation was devoted to the prediction of school and job performance of machinist apprentices.

QUALITY CONTROL WORKSHOP

DEVELOPING PERFORMANCE TESTS FOR TRAINING EVALUATION: A JOB AID FOR TEST DEVELOPERS

By WILLIAM C. OSBORN
HumRRO Division No. 2
Fort Knox, Kentucky

A performance test is a template -- a template modeled from a job task and used to gauge the similarity of a trained behavior to that job task. This view of performance tests implies a straight-forward approach to their development. One simply recreates the circumstances of the job task, asks the trainee to perform the task, and then records that he did or did not do it. Unfortunately, from our own experience we know that it is not this simple. Many practical problems intervene to complicate the process. We often find that a job has so many tasks that days would be needed to test on them all. Occasionally the equipment, terrain and other support requirements prohibit a realistic test for even a single task. At other times we run into standards of task performance that are difficult to translate into a pass-fail criterion for scoring. We also have found that trainers need more than pass-fail results, they need diagnostic information to tell them why their trainees failed, if they did.

These are some of the major problems encountered by test developers, though by no means all. For the most part, the kinds of test development problems that we encounter in the field of training evaluation are not the same as those encountered in the field of aptitude testing. Thus we have found the traditional body of academic literature on test development to be poorly suited to our needs. Certainly the basic notions of reliability and validity apply to any test development effort, but in our field the exotic, sophisticated formulae which fill most books on test development are of little use.

What is vitally needed in the field of training evaluation, it seems to me, is a how-to-do-it manual for test developers -- one that responds to the variety of practical constraints and problems that occur in the process of constructing tests for the myriad tasks spanned by some eight or nine hundred Army jobs. I sincerely wish that I had such a manual for you today. But I don't.

What I do have is intended to be a very small step in that direction. I have attempted to chart the major action points in the course of developing a test for training evaluation. You each have a copy of this chart, and it is my hope that you will find it a useful framework for discussing the problems and practices of test development.

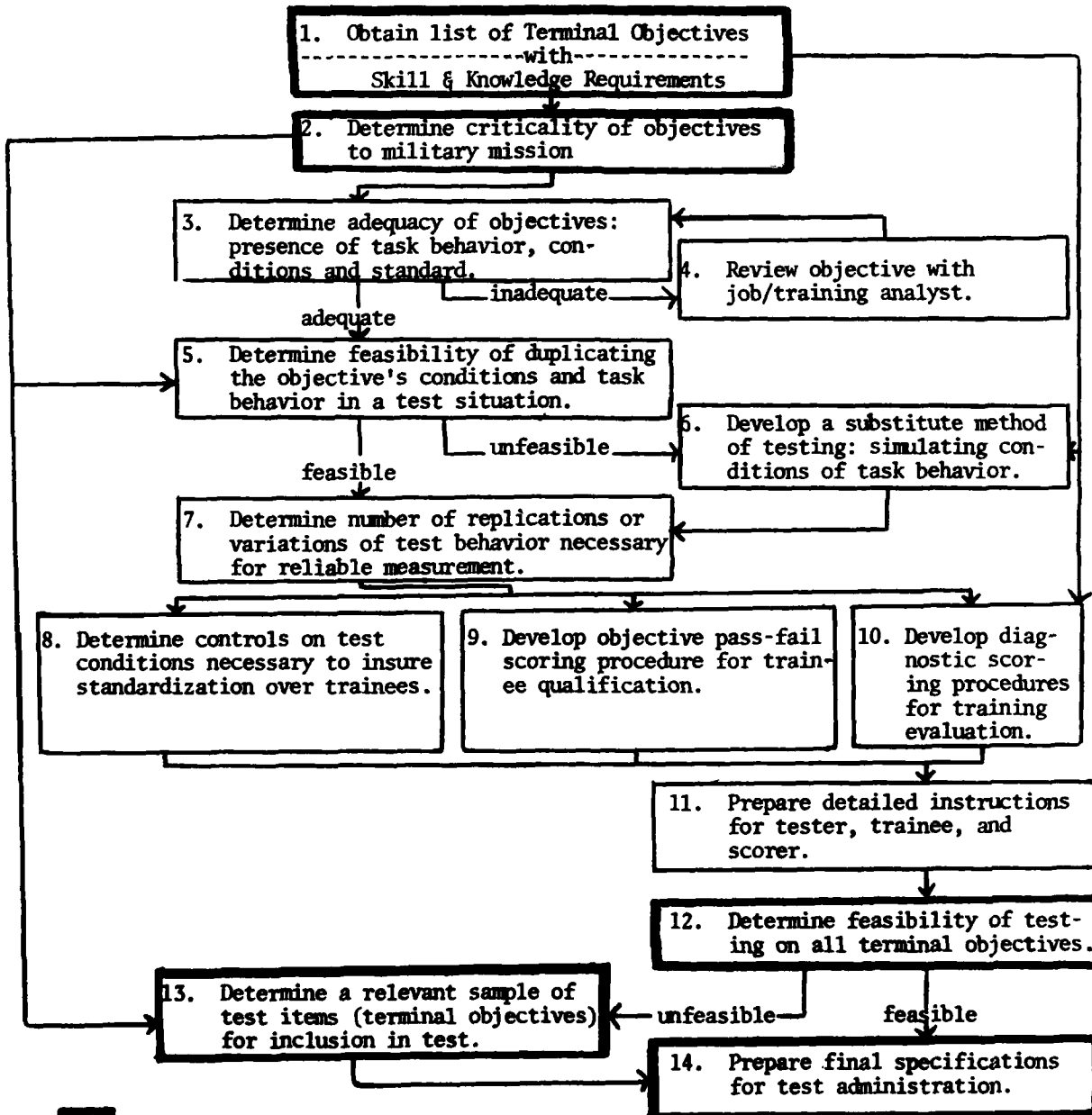
Before going any further, however, there are two matters of terminology which need clarification. The first has to do with the concept of performance testing. I choose to use this concept (at least today) to designate the test or tests, normally developed and administered by a quality control agency on completion of training for the two explicit purposes of qualifying trainees and evaluating training. This is to be distinguished from the development and use of test by trainers for monitoring student progress within and between stages of training. Secondly, you will find that I use the term test item in referring to the evaluation of behavior involved in a single job task, and the term test in referring to the aggregate of these items over an entire job or job sector purportedly covered by the training program. I am not asking you to buy these labels, but to bear them in mind for the moment.

Now let us return to the process of test development as outlined on the chart. What I would like to do is to proceed through these 14 steps, one by one, giving you a brief summary of my thoughts on the 'why, what and how' of each. Then I would like to hear your thoughts, experiences, complaints and suggestions regarding the various aspects of test development as outlined.

The first three steps on the chart refer to assembling information that should routinely be supplied to the test developer. He should only have to verify its presence and completeness, and not make judgements about its accuracy. As stated in the first step, test development begins with the objectives for the job or job sector for which people are to be trained. These are sometimes termed job objectives -- more often, terminal training objectives. Whatever they are called they are the master list of specifications derived from the job, and from which both training developers and performance test developers separately begin their work. As test developers, our goal is to develop a performance test item for each and every objective; though this is not to imply that our final test will necessarily encompass all objectives. In addition, each objective should be accompanied by a supporting list of skill and knowledge requirements which are used in later stages of test development.

The information designated in Step 2 should also be available as a matter of course. The relative importance of each objective

STEPS IN PERFORMANCE TEST DEVELOPMENT



☒ Steps pertaining to development of total test

☐ Steps pertaining to Development of each test item

as judged in terms of mission capability, represents data that is necessary in making trade-offs later in the test development process.

Step 3 suggests that each objective must be reviewed to make sure it is all there. We all know that, in addition to a stated task behavior, an objective should contain stated conditions and standards of performance. If any of the three elements is missing, or if any are unclear to the test developer, he should get together with the task analyst and, as indicated in Step 4, obtain a clear statement of the missing or confusing element. Performance standards are the most common source of trouble, and if a fair and meaningful pass-fail criterion is to be established for a test item, the developer must have an unequivocal standard of task performance to work from.

In Step 5, test item development really begins. Here the developer must judge the feasibility of duplicating in a test situation the conditions and behavior called for in the objective. Normally, of course, our view is that well stated objectives are blueprints for testing -- in fact, dictating what the test conditions will be. But occasionally we encounter an objective calling for the use of job relevant equipment, terrain, support personnel, or a time frame which exceeds the resources available to the test agency. In these instances the developer must carefully weigh the criticality of the objective (from Step 2) against the cost factors before deciding that full realism cannot be afforded, because invariably some degree of relevance is lost as we depart from the test specifications given in the objective.

When it is decided that the conditions of the objective cannot be duplicated in the test situation, a substitute technique must be developed, as indicated in Step 6. This is perhaps the most subtle and challenging aspect of the development process. Here a developer's inventiveness is often called for in devising a method and conditions for testing that will call for the demonstration of a behavior that is as similar as possible to the behavior stated in the objective. Too often in this situation developers resort to paper-and-pencil tests measuring knowledge of the task, and approach that in most cases can be safely rejected out of hand. In considering simulation options developers have a useful check available in the task's skill and knowledge requirements. The relevance of a proposed test method may be evaluated by checking the number of skill and knowledge components of the task that are called for in the method.

Once a general method of testing is determined with Step 5 or Step 6, the developer turns his attention to the matter of achieving

measurement reliability. In Step 7 he must again look at the objective in terms of repetitions or variations of the behavior implied. In most cases this will be explicitly given. For a specific skill such as disassembling a rifle or installing a carburetor a single demonstration of the behavior is all that is normally called for. On occasion, however, with generalized skills or generalized behaviors, the number of repetitions of the behavior may or may not be clearly stated in the objective. An objective specifying that something will be done correctly 9 out of 10 times creates no problem for the test item developer, as 10 repetitions are required. On the other hand the standard may be phrased in terms of correct performance on 90% of the trials. Here a decision must be reached on an appropriate number of repetitions of the performance to ask for in the test item. More generally, the important consideration in Step 7 is whether a large enough sample of trainee performance is being required so that success or failure does not result largely from chance. Here again the test developer must make some trade-off between time or cost factors and reliability of the measured behavior.

Step 8 pertains to another aspect of test reliability -- the standardization of the conditions under which a test item is administered. Here the important factors are the instructions and environmental conditions under which the test item is given. Instructions should be identical for everyone. They should be clearly and simply stated, leaving nothing to the interpretation or misinterpretation of the trainees taking the test. Things such as the method of scoring and whether speed or accuracy is important should be stressed in the instructions. Also conditions pertaining to test supplies and environmental factors should be constant for all personnel. Items of equipment worked with or on during testing should be restored to their pretest condition if they are used by successive trainees. Similarly, environmental factors such as visibility, temperature, attitude of the tester, time of day, and the like must be stabilized.

In Step 9 a final aspect of measurement reliability is considered. Here procedures for translating an observed trainee performance into a pass-fail score must be developed. Provision for this type of scoring should be structured so that only the more reliable human skills are used. That is, the scoring activity should be reduced to one of matching or comparing the test item response with some model of the acceptable response. If the model response on a test of rifle marksmanship is defined as a hole in the bullseye, then the scorer has a relatively easy task in judging the acceptability of the response made by the rifleman. Unfortunately, responses for many test items cannot be judged in this "either-or" fashion, but require a "more-or-less" type of judgement. In these cases the developer should always strive to break down the model response

into elements so that comparative judgements can be made more easily by the scorer. This may often entail preparing a check-list of the necessary components or features of the model response.

In Step 10 a supplementary scoring procedure is developed for use in diagnosing reasons for trainee failure on the test item. Pass-fail scoring is sufficient in meeting the primary mission of quality control, which is the certification of trainee job readiness. However, the secondary mission, that of training program evaluation, is best carried out by providing the trainers not only with the incidence of pass and failure for an objective, but also feedback on why trainees failed. One way to obtain this data is through a check-list developed from the skill and knowledge requirements of the task -- a check-list to be used by the tester in recording why the trainee failed a test item. When accumulated over a number of test item administrations, this diagnostic information will normally provide a stable picture of the reasons for failure which trainers may then use to selectively revise and strengthen their program.

In Step 11 the test developer simply brings together the products of previous steps and formats the final test item. Detailed instructions to the tester covering test materials, equipment, procedures, precautions, etc., are spelled out. Also the directions to be read to the trainee by the tester, and the scoring procedure should be written out.

The final three steps on the chart pertain to assembly and administration of the final form of the test. In Step 12 a decision is made on whether time permits testing on all objectives -- that is, administration of all test items. If it is not feasible to do so, an appropriate sample of test items has to be selected (Step 13). As indicated in this step, the main criterion for sampling should derive from criticality ratings of the objectives. An exact procedure for doing this will depend on the categories originally used for reporting criticality. Generally, the developer would first include all "essential" or highly critical items, and then sample from the remaining. Wherever sampling is necessary the usual practice is to vary the sample from one administration to the next so that all test items are used sooner or later. Variations in the sample should not be systematic in the sense that trainers or trainees can anticipate what items are going to appear.

In Step 14 final guidance for test administration is prepared. Training for testers may have to be designed; list of equipment and materials prepared; and scheduling worked out. If testing is to be done individually, it is usually a good idea to prescribe a "county-fair" layout of test situations. This serves purposes of economy, as well as permitting test items to be administered in

varying order. In addition, security precautions must be specified to insure, for example, that one trainee cannot benefit by observing another's performance, or that trainees do not talk among themselves during test administration.

This completes my account of performance test development. Now I would like to hear your views and comments. Let me open the discussion by asking this: If the steps in this chart were elaborated by including procedures and examples, do you think it would be a useful job aid to performance-test developers in the Army?

QUALITY CONTROL WORKSHOP

DESIGN OF OBJECTIVELY SCORED PERFORMANCE TESTS AND USE OF TEST FEEDBACK

By ROBERT N. JOHNSON, GS-13
Asst Director of Instruction, USAAGS

I. INTRODUCTION.

Mr. Osborn has just presented an overview of a model for the development of performance tests which should materially contribute to the establishment of an effective Training Quality Control system. I'm sure that most of us will agree with the overall model presented although we may have reservations concerning specific aspects thereof. The major problem area, as specified by Mr. Osborn, is the absence of a "how-to-do-it" manual for use at all CONARC Schools. The need for a manual, of course, presupposes a prior discussion of several basic issues. Issues like "Could such a system actually be implemented in an operating School?", "How much work would it actually entail?", "What are the problem areas?", and "Is it worth the effort?"

At this time I'd like to present a more detailed view of a system which fits within the framework of the model described. This system has been in actual operation at the US Army AG School for several years. As Mr. Osborn indicated, there are two primary objectives which a performance test can serve. One, to validate the job readiness of the graduate and two, to provide feedback to the instructional system itself. The AG School system emphasizes the second objective. It does not exactly match the 14 steps outlined but the purpose, logic, general sequence and results are almost identical. By first presenting an overview and then following with an operating example, we feel that the problem areas for discussion become more readily apparent. After this presentation we will take a 15 minute break and open the workshop for discussion when we return. Now for the operating example.

II. OBJECTIVES OF PERFORMANCE TESTING.

What should be the objectives of a good test? To whom should it provide a service? What functions should it perform? This transparency outlines one approach to what a good test should accomplish. (Incl 1 VG-Objectives)

a. In order to serve as a valid quality control device, test results must provide detailed feedback which will facilitate the

continuous upgrading of our instruction. If we can pinpoint specific weaknesses, in specific portions of our lesson plans, upgrading of the instruction is simplified. In one sense then, we can say that the tests should serve the instructional system to include the instructor and instructional managers. In another sense however, we can say the test should serve future graduates in that it should provide for better and better instruction as time goes on.

b. A valid test must also identify individual student weaknesses in order that appropriate remedial instruction can be provided. In this sense the test should serve the individual student and his instructor.

c. By validating individual student ability to perform field tasks to field standards, the test is serving the commander in the field by reducing his training load and the OJT time necessary for the graduate to achieve job competency.

III. DESIGN CONSIDERATIONS.

What then, are the design considerations which will facilitate the attainment of these objectives? (Incl 2 VG-Design Considerations)
In view of Mr. Osborn's presentation I will not take the time to discuss each of these design considerations in detail but I would like to emphasize a few points.

a. Too often tests are designed to measure whether or not the student has learned what the instructor taught him? This is measurement of the student only! We are trying to measure the effectiveness of the instructional system in addition to the student. Design of performance tests directly from the objective, prior to and independent of instructional design, results in measurement of the instruction itself. If the overwhelming majority of students can achieve the established standard, the instruction has been successful, if not the instruction must be modified.

b. Knowledge for knowledge's sake is useless. Valid tests must not only measure student knowledge but, in addition, must measure whether or not the student can apply the knowledge in the appropriate situation, at the right time, in the appropriate sequence and in combination with other knowledges. A test which requires the student to actually perform objective tasks to establish standards will satisfy all these requirements.

c. Realistic Approximation.

(1) The validity of a performance test is measured in terms of the degree to which it represents reality. As I will show later the

more realistic the test instrument, the more instructional feedback it produces. Within cost and time limitations, the performance test must provide conditions and cues identical to those which exist on the job.

(2) Actual job tasks are always performed in a specific environment with existing conditions and limitations. The unit to which assigned, its mission and echelon, existing local policies and procedures, resource limitations, time, data and place all interact to influence what is considered to be adequate performance of a task on the job. The appropriate aspects of the environment must be included in the test conditions.

(3) Options available to the student should also be identical to those available in a real world situation. This design consideration eliminates the use of multiple choice questions for the majority of performance tests. A multiple choice question identifies a problem and then give the student four or five possible answers to select from. In the real world, however, the student is usually required to determine whether or not a problem exists, determine for himself the available options and then select the appropriate option. To identify the problem for the student and to offer alternatives is not realistic and therefore reduces the effectiveness of the feedback.

d. Machine Processing of Test Results. In a single 5 week course of instruction we administer 6 tests with a total of 915 scoreable items for each student. With a student input of 3,700 for the current FY, we will have to manipulate over 3 1/2 million items of data. Considering that each piece of data is handled at least once for item analysis purposes, and once for grading purposes, we have over 7 million manipulations for a single course of instruction. We estimate that we handle over 25 million data manipulations a year for all of our 30 courses. It is obvious that machine processing of test results is an absolute necessity in an effective Quality Control System.

IV. TYPES OF PERFORMANCE TESTS.

Now, lets talk a little about performance tests themselves. Performance tests can be classified many ways. We recognize three types of performance tests as shown here. (Incl 3 VG-Types of Performance Tests)

a. The Answered Scored PT is an instrument on which the student records his answers and the answers are scored directly by a machine. Most of you may consider this to be a paper and pencil test rather than a performance test but I will show you, later, how a paper and pencil test can actually function as a performance test. The main advantages of the Answer Scored PT is that it does not require the time of a qualified instructor to administer or to score and that it can be group administered. Group administration, of course, reduces POI time required for testing.

b. In the Product Scored PT, the student produces a tangible product which is later scored by an instructor. The advantage here, is that even though the instructor has to score the product, he need not be involved in the test administration. Scoring can be accomplished later at the convenience of the instructor. Product Scored tests can normally be group administered.

c. The Observer Scored PT requires a qualified instructor to observe the actual performance of the student in order to score the test. This is obviously the most restrictive of the three types in terms of instructor resources.

V. DESIGN OF ANSWERED SCORED PERFORMANCE TESTS.

At this point I'd like to show you an example of our approach to the development of an actual performance test based upon the objectives and design considerations I have just outlined. Although we do have many product scored and observer scored performance tests at our school I chose not to use either as my example today. I'm sure most of you are involved with "hard skills" and therefore are already familiar with the design of product and observer scored performance tests. At the AG School, however, we are concerned with many soft skills and have found that the Answer Scored Performance Test adequately meets our needs. I will therefore illustrate this type of test in hopes that it may give you some new ideas on how to approach the soft skills especially in officer, supervisory or management areas. I'd like to use our Army Recruiter Course. I am sure you are all generally familiar with the mission of a Recruiter - to secure enlistments in Regular Army. In effect the recruiter is a salesman whose product is the US Army. After following up leads, he finds a face to talk to, and then gives a sales pitch to interest the prospect. Once the prospect is interested, his next task is to determine the prospect's eligibility for enlistment. It is this task of "determining eligibility" that I will use as an example. The Army Regulation (AR 601-200) concerning eligibility has over 180 pages including 8 changes to the basic regulation. The regulation includes an enormous amount of narrative instructions and numerous charts and tables. A recruiter must check out and document the eligibility of each prospect under this regulation. Eligibility is based upon many factors including age, moral, mental and physical qualifications, education, marital status, etc.

As indicated earlier the test should be designed directly from the analysis of the task, prior to and independent of instructional design. During the task analysis portion of the Systems Engineering effort in this course, two logic trees were prepared; one for prior service applicants, and one for non-prior service applicants. I'd like to show you how we use these logic trees as a test outline. This is the logic

tree for determining eligibility for non-prior service personnel (Incl 4 LT-Determining Eligibility). You will note that on the face of the tree the task, conditions and cues are described, also the references used to prepare this particular tree. If you will open the paper you will see the mental decisions and physical actions necessary to perform this specific task. The logic tree is nothing more than a schematic, outlining how a task is performed. The tree starts with an oval in the upper left hand corner. From that point you merely follow the arrows. Questions are contained in hexagons and are always answered with a "yes" or a "no". Actions are in rectangles. Circles permit you to skip portions of the procedures.

The first question we must face is which of the three types of performance tests is appropriate for use with this task? Although we prefer performance tests in the priority shown earlier for reasons of economy the types of performance test to be selected depends upon the nature of the task itself. You will find as your next handout another logic tree called "Determining the Appropriate Type of Performance Test". (Incl 5 LT-Determining Type of PT). This task, of course, is performed by the training analyst or test designer. This logic tree has been written to insure that you will select the most economical type of performance test for use with any specified task.

(Lead the group through the logic tree using the task of "Determining Eligibility for Enlistment. Blocks 4, 5, 6, 7, 8, 9, 10 and 11). As you can see the Answer Scored Performance Test, our most economical test in terms of resources, is appropriate for use in evaluating this task.

Our next step is to design the answer sheet. Note that we design the answer sheet before we design the test instrument itself. In order to design the answer sheet we must again examine the logic tree for "Determining Eligibility for Enlistment". (Incl 4). Since the logic tree includes all the variations of this task that can occur in the real world we can use the tree itself as a test outline. First off, we examine the tree and determine precisely what is involved in the task and what the expected results of adequate performance are. In this case we find that the student is examining a series of factors to determine three things (1) the availability of adequate documentation, (2) the need for waiver of certain disqualifications or (3) total ineligibility. Accordingly we design the answer sheet as shown here. (Incl 6 VG-Answer Sheet). In effect we have a matrix answer sheet in which each factor under consideration is listed down the left column and all possible responses are listed across the top. Note that in this case it is quite possible that there may not be anything required on one or more of the factors. We could have added another column to indicate "no problem" but in this case we chose not to. These are not multiple choice questions but rather a categorization of the many possible responses possible on the job in the field.

We now have an answer sheet. Our next step is to design the test instrument itself. The conditions and cues on the face of the logic tree outline what must be provided in the test instrument. In this case they are the results of the interview, documentation provided by the applicant, and documents from other sources. We must now create enough situations to include in the test to provide an adequate sampling of the students ability to perform the whole task. Here again, we use the logic tree as a test outline. By tracing a single path through the logic tree we have a blue print outlining one situation to be included in the test instrument. Other situations are then traced to cover the remaining blocks of the tree. In order to give adequate coverage of this task we came up with the requirement for 18 situations and divided them into 3 versions of the test with 6 situations each. We alternate test versions to individual classes but the students are taught so that they can pass any one of the three versions.

I'd like now to show you a situation out of the A Version of this test (Incl 7 Test Situation). This situation portrays a non-prior service female named Susan Settles. The information shown on the Prospect Card was secured by the recruiter during the interview. The only documentation produced by the applicant is a birth certificate and a drivers license. The details of this situation match exactly the path we just traced through the logic tree. The student examines the documentation presented and makes the same determinations that he must make on the job. The only difference is the added requirement that he record these decisions on the answer sheet. There are no real questions on the test. The student must determine if there is a problem and how to resolve the existing problem. The test itself does not give him a clue as to the existence of the problem. There may be a problem concerning moral qualifications or there may be none. The instructions for the test include environmental conditions vital to proper accomplishment of the task, who he is, where he is, date, time and background material. If you will look at the other 3 situations on the supplement you will see that each is slightly different. Situation 2 concerns a prior service male and therefore includes a copy of his discharge papers. Each case was developed by using the logic tree as a test outline. But in each case the answer sheet is identical.

Once the test is completed it is first administered to experienced members of the instructional staff. This helps to debug the instrument and to give us an estimate of the amount of time to allow for administration. In addition it gives us a handle on the appropriate P/F point. From experience over the last 5 years we feel that a well designed performance test plus the critique, will take about 10% of the POI time that it takes to teach the task (20 hours for instruction- 2 hours for testing).

The test is now administered to a group of students on a trial basis to further debug the instrument. If possible the test should be administered to a group which has completed the pre-systems engineering (knowledge type) instruction. This gives the added advantage of developing pre-systems engineering data for later comparison.

Once the test has been debugged, all test versions plus the logic tree (or TAIS) are turned over to the instructional department with instructions to teach the task as represented by the logic tree so that the students can pass any versions of the test. In addition to the three test versions, one additional version is developed as a practice test. This is used in the classroom as a practical exercise to insure that the students understand the mechanics of taking the test. It also serves as an excellent review.

VI. USE OF INSTRUCTIONAL FEEDBACK.

Instruction is now conducted followed by a practice test and critique, then one version of the actual test is administered. Students mark their responses on the answer sheet printed on the test supplement, in addition they punch their responses on a punch card. Punch cards are collected immediately following the test but the test supplements remain with the students for use during the final critique.

Punch cards are run through the computer and the following feedback is produced. (Incl 8 VG-Item Analysis) This printout contains the item analysis in terms of miss rates. Each item in the test which has a miss rate of over 10% is triple asterisked by the computer. This printout is designed to furnish feedback to the instructor. 10% or less on any element of performance is acceptable. Anything over 10% indicates a problem area. I'd like now to concentrate on the use of the item analysis as a means of instructional feedback and I'd like to show you some comparisons of the feedback which pertain to the case of Susan Settles which we designed earlier. (Incl 9 VG-Eligibility). The four columns show the miss rates in 4 of the 8 classes out of the first cycle of instruction in July and August of this year. It should be noted that since these 8 classes were conducted simultaneously our instructional feedback mechanism was not in operation. I think the figures illustrate item difficulty and item reliability rather clearly. Items not shown had less than a 10% miss rate in all classes.

After instruction in the first cycle, each individual instructor received the item analysis data shown on the printouts. I'd like now to just show you one or two examples of how the instructor uses the item analysis to improve his instruction. Lets look at Item 1 in the analysis. This item concerns problem areas with the name of the applicant and has a high miss rate through all classes. If you will

look again at the actual test situation (Incl 7), you will see that the girl was born with the name Susan Settles as documented by her birth certificate and wants to enlist under that name. The information from the interview as shown on the prospect card shows she was married to and divorced from a Charles P. Martin. Her drivers license shows her name to be Susan Martin. The problem as shown on the logic tree is to determine whether or not she is enlisting under her legal name. The documentation displayed on the test supplement is inadequate to resolve this question. If her divorce decree restored her maiden name of Settles, there is no problem. If it did not, the recruiter must prepare and Susan must sign, a DA Form 3784-R (Statement of Name Change). In either event there is additional documentation required before this girl can be enlisted (the form or the divorce decree). The proper answer therefore is No. 1 - Additional documentation required. With 8 different instructors there were many reasons why the miss rate was so high. In most cases we found that the instructors were teaching male eligibility only. The point, however, is that once the problem area is clearly identified improvement of instruction becomes possible. In Item 2, Age, the only problem is that the girl is under 21 and needs parental consent. A male however need only be 18 years of age. In most cases the instructors failed to cover age requirements for females. With respect to Item 20, a divorce certificate is required to prove she is no longer married because married women cannot enlist unless they have prior military service. By comparing what was taught with the test results, instructional problems can be isolated and eliminated.

Now lets look at the test results from the second cycle of classes. Although we used the A version of the test for all eight classes in the first cycle we only used it for only four classes in the second cycle. Here are the results for two of those 4 classes from cycle 2 compared with the classes taught by the same two instructors in cycle 1. (Incl 10-VG Miss Rates by Instructors). Obviously both instructors achieved major gains in student performance. Both instructors, incidently, gave the test feedback credit for isolating their problems for them. You will find a logic tree outlining the task of "Using item analysis results to improve instruction" among your handouts (Incl 11). You will also find a research study covering the use of the logic tree as a job aid *(Incl 12). These documents are for your later perusal.

I've spent considerable time discussing the design of Answer Scored Performance Tests and the use of instructional feedback of these tests. I have emphasized this type of test for two reasons. First, of all because it is the most difficult and time consuming type of test to design. And secondly, because I want to debunk the commonly accepted

*Omitted. Copy of Incl 12 is obtainable from: Educational Advisor
ATSAG-EA
USArmy Adj Gen Sch

view that a paper and pencil test cannot be a performance test. If the task itself is essentially a paper and pencil exercise or if the task is essentially a mental exercise it is possible to design a paper and pencil performance test.

VII. DESIGN OF PRODUCT AND OBSERVER SCORED PERFORMANCE TESTS.

Our approach to the design of tests which score the product of student performance is probably not much different from that used in other CONARC Schools. The probable differences are that we again use the logic tree as our test outline, and that we design the faculty produced score sheet in a manner which can be computer scored to produce feedback which isolates problem areas. I like to just take a minute to show an example.

In our postal operations course we teach students to issue money orders, cash money orders and make the daily business report. The test requires the student to actually cash 5 money orders, issue 5 money orders and make a business report. In this test, the product of performance, that is the money order or business report itself, is scored item by item on a go/no go basis by a knowledgeable instructor. Results are punched into a card and run through the computer. This viewgraph shows the results over a two year period. (Incl 13 VG-Cash Money Orders). Note the substantial improvement between class 3 and class 4 of 1969. This is a normal result of the use of feedback, the largest percentage gain is evidenced in the first few classes. Note, however, that two years later the quality of performance is substantially the same but look at the reduction in POI hours. The key point here is that once we have improved our instructional design to a point which consistently produces the desired quality of student performance, we immediately shift our attention to the question of cost effectiveness. The question is "are we getting good results because we have dedicated too much of our resources to the instruction of this task". We attempt to answer this question by reducing resources and watching the effect on student performance in terms of test results. Early in FY 1970 we started reducing the number of POI hours dedicated to instruction in money orders, by mid 1970 we had reduced it to 21 hours. This year we reduced it to 16 hours with no material change in student performance. In this instance test results serve a function in resource allocation as well as in our quality control.

Let us now take a quick look at objective type tests which require actual observation of student performance in order to complete the score sheet. For example, one task of a punch card machine operator is to sort a deck of cards utilizing a sorting machine. There is a product, in the sense that a deck of cards is produced, but examination of the final deck does not adequately identify the student errors which

may result in an erroneous product. According we give the instructor a score sheet on which he scores each step in the procedure on a go/no go basis as shown here (Incl 14 VG-Sorter). Again we process the results through the computer. I won't show you any examples of results since the results are similar to the Money Order results you've just seen.

VIII. CLOSE.

This has been a rather detailed view of an operating system designed for the specific purpose of providing detailed feedback which facilitates the continuous upgrading of the instruction itself. Please note that this presentation has been restricted to test design considerations for tasks which can be objectively scored. After the break we will open the subject for discussion and then get into what appears to be a major problem at many schools - how to handle tasks which cannot be objectively scored. Lets take a 15 minute break.

THE FOLLOWING IS A SUMMARY OF THE DISCUSSION WHICH FOLLOWED THE ABOVE PRESENTATION.

- a. It was generally accepted that performance test should accomplish two objectives. One to establish the job readiness of the student (Pass/Fail on each task), and two, to identify areas of instructional and student weakness (feedback). Although most schools are designing tests to accomplish the first objective, few had expended any effort toward the development of an instructional feedback loop. The model presented by the AG School was accepted by most as a excellent approach to development of useful feedback, but several doubts were raised as to the validity of using the same data to determine job readiness.
- b. It was generally agreed that the materials available to a student during a test should be identical to those available on the job even though additional materials might be used during the instructional process.
- c. Most agreed that the multiple-choice and other knowledge type test formats were inappropriate for use in performance tests. Most schools also found the conventional item discrimination indices were of little value.
- d. The need for some type of ADP support for the Quality Control effort was agreed upon. The availability at each school of \$10,000 per annum which could be used for this purpose was noted.

PERFORMANCE TESTING OBJECTIVES

1. TO PROVIDE AN INTERNAL FEEDBACK WHICH WILL FACILITATE THE CONTINUOUS UPGRADING OF THE INSTRUCTION AND THUS THE QUALITY OF THE GRADUATE.
2. TO IDENTIFY SPECIFIC STUDENT WEAKNESSES WHICH REQUIRE REMEDIAL INSTRUCTION.
3. TO INSURE THAT ONLY THOSE STUDENTS WHO CAN ADEQUATELY PERFORM THE TRAINING OBJECTIVES ARE PERMITTED TO GRADUATE.

PERFORMANCE TEST DESIGN CONSIDERATIONS

1. DESIGNED PRIOR TO AND INDEPENDENT OF INSTRUCTIONAL DESIGN.
2. MEASURES STUDENT ABILITY TO PERFORM OBJECTIVE TASKS RATHER THAN MASTERY OF KNOWLEDGE OR SKILL COMPONENTS.
3. REALISTIC APPROXIMATION
 - A. IDENTICAL CONDITIONS & CUES.
 - B. OPTIONS SAME AS ON THE JOB.
4. OBJECTIVE SCORING (GO/NO GO).
5. CRITERION REFERENCED GRADING (PASS/FAIL).
6. MULTIPLE HURDLE CONCEPT.
7. MACHINE PROCESSING OF TEST RESULTS.

PERFORMANCE TESTS

<u>TYPES</u>	<u>INSTRUCTOR INVOLVEMENT</u> <u>ADMIN</u>	<u>SCORING</u>	<u>GROUP ADMINISTERED</u>
1. ANSWER SCORED PT	NO	NO	YES
2. PRODUCT SCORED PT	NO	ANY TIME	YES
3. OBSERVER SCORED PT	YES	POI TIME	NO

UNITED STATES ARMY ADJUTANT GENERAL SCHOOL



LOGIC TREE FOR

DETERMINING ELIGIBILITY FOR NON PRIOR-SERVICE ENLISTMENT (ARCC 2-3-11)

TASK: To determine eligibility for enlistment of non-prior service personnel under AR 601-210

CUE: Willingness of a prospect to initiate processing

CONDITIONS: To be conducted during or subsequent to a qualifications interview during which information shown on the Prospect Card (USAREC Form 200) was elicited from the prospect. Access to documentation provided by the prospect, AR 601-210, appropriate US Government, DD, DA and USAREC Forms and enlistment screening tests.

Source Data
AR 601-210 w/C8
Local SOP

ARCC LT 2-3-11



FORT BENJAMIN HARRISON, INDIANA 46216

INCL 4

204

UNITED STATES ARMY ADJUTANT GENERAL SCHOOL



SYSTEMS ENGINEERING OF TRAINING

LOGIC TREE FOR

DETERMINING TYPE OF PERFORMANCE TEST (SET 5-1)

JOB: Training Analyst/Test Designer

TASK: To determine the appropriate type of performance test

CUES: Decision to test a task selected for training

CONDITIONS: Availability of a logic tree (LT) or task analysis
information sheet (TAIS)

Prepared for the Quality Control portion of the CONARC Training
Workshop, OCT 5-7, 1971. (To be incorporated into the USAAGS
Instructor Training Course)

Comments and/or suggestions should be forwarded to USAAGS, ATTN:
Director of Instruction. Telephone - Autovon 699-3648 or 699-
3703; civilian (317) 542-3648 or 542-3703

Source

SET 5-1
AUG 71

CON Reg 350-100-1 (SET) dtd 1 Feb 68 w/C1
USAAGS Reg 350-100 (SET) dtd 1 Jul 68 w/C1
USAAGS Reg 350-2 (Resident Student Evaluation
Program dtd 30 Jan 69
USAAGS Local SOP



FORT BENJAMIN HARRISON, INDIANA 46216

Incl 5

<u>QUALIFICATION CATEGORIES</u>		<u>FURTHER DOCUMENTATION REQUIRED</u>	<u>WAIVABLE DISQUALIFICATION</u>	<u>NOT ELIGIBLE AT THE PRESENT TIME</u>
NAME	1		X	X
AGE	2		3	4
CITIZENSHIP	5		X	6
SOCIAL SECURITY ACCOUNT NUMBER	7		X	X
MORAL QUALIFICATION	8		9	10
TEST SCORES	11		12	13
PHYSICAL QUALIFICATIONS	14		15	16
CIVILIAN EDUCATION	17		18	19
MARITAL/DEPENDENCY STATUS	20		21	22
DRAFT STATUS	23		X	24
RESERVE STATUS	25		X	26
PRIOR MILITARY SERVICE	27		28	29
REFERENCES (RECOMMENDATIONS)	30		X	X

Incl 6

24.8

TODAY IS 1 MARCH 1974

SETTLES, SUSAN PHYLIS				5216 N. Boy Scout Road				S46-0095			
FIRST NAME				2. ADDRESS				3. TELEPHONE			
1. LAST NAME				4. BIRTH DATE				5. SEX			
A. BIRTH DATE				B. BIRTH DATE				C. BIRTH DATE			
D. BIRTH DATE				E. BIRTH DATE				F. BIRTH DATE			
G. BIRTH DATE				H. BIRTH DATE				I. BIRTH DATE			
J. BIRTH DATE				K. BIRTH DATE				L. BIRTH DATE			
M. BIRTH DATE				N. BIRTH DATE				O. BIRTH DATE			
P. BIRTH DATE				Q. BIRTH DATE				R. BIRTH DATE			
S. BIRTH DATE				T. BIRTH DATE				U. BIRTH DATE			
V. BIRTH DATE				W. BIRTH DATE				X. BIRTH DATE			
Y. BIRTH DATE				Z. BIRTH DATE				AA. BIRTH DATE			
AB. BIRTH DATE				AC. BIRTH DATE				AD. BIRTH DATE			
AE. BIRTH DATE				AF. BIRTH DATE				AG. BIRTH DATE			
AH. BIRTH DATE				AI. BIRTH DATE				AJ. BIRTH DATE			
AK. BIRTH DATE				AL. BIRTH DATE				AM. BIRTH DATE			
AN. BIRTH DATE				AO. BIRTH DATE				AP. BIRTH DATE			
AQ. BIRTH DATE				AR. BIRTH DATE				AS. BIRTH DATE			
AT. BIRTH DATE				AU. BIRTH DATE				AV. BIRTH DATE			
AW. BIRTH DATE				AX. BIRTH DATE				AY. BIRTH DATE			
AZ. BIRTH DATE				BA. BIRTH DATE				BB. BIRTH DATE			
BC. BIRTH DATE				BD. BIRTH DATE				BE. BIRTH DATE			
BF. BIRTH DATE				BG. BIRTH DATE				BH. BIRTH DATE			
BI. BIRTH DATE				BJ. BIRTH DATE				BK. BIRTH DATE			
BL. BIRTH DATE				BM. BIRTH DATE				BN. BIRTH DATE			
BO. BIRTH DATE				BP. BIRTH DATE				BQ. BIRTH DATE			
BR. BIRTH DATE				BS. BIRTH DATE				BT. BIRTH DATE			
BU. BIRTH DATE				BV. BIRTH DATE				BW. BIRTH DATE			
BX. BIRTH DATE				BY. BIRTH DATE				BZ. BIRTH DATE			
CA. BIRTH DATE				CB. BIRTH DATE				CC. BIRTH DATE			
CD. BIRTH DATE				CE. BIRTH DATE				CF. BIRTH DATE			
CG. BIRTH DATE				CH. BIRTH DATE				CI. BIRTH DATE			
CJ. BIRTH DATE				CK. BIRTH DATE				CL. BIRTH DATE			
CM. BIRTH DATE				CN. BIRTH DATE				CO. BIRTH DATE			
CP. BIRTH DATE				CQ. BIRTH DATE				CR. BIRTH DATE			
CS. BIRTH DATE				CT. BIRTH DATE				CU. BIRTH DATE			
CV. BIRTH DATE				CW. BIRTH DATE				CX. BIRTH DATE			
CY. BIRTH DATE				CZ. BIRTH DATE				CA. BIRTH DATE			
CB. BIRTH DATE				CC. BIRTH DATE				CD. BIRTH DATE			
CE. BIRTH DATE				CE. BIRTH DATE				CE. BIRTH DATE			
CF. BIRTH DATE				CF. BIRTH DATE				CF. BIRTH DATE			
CG. BIRTH DATE				CG. BIRTH DATE				CG. BIRTH DATE			
CH. BIRTH DATE				CH. BIRTH DATE				CH. BIRTH DATE			
CI. BIRTH DATE				CI. BIRTH DATE				CI. BIRTH DATE			
CJ. BIRTH DATE				CJ. BIRTH DATE				CJ. BIRTH DATE			
CK. BIRTH DATE				CK. BIRTH DATE				CK. BIRTH DATE			
CL. BIRTH DATE				CL. BIRTH DATE				CL. BIRTH DATE			
CM. BIRTH DATE				CM. BIRTH DATE				CM. BIRTH DATE			
CN. BIRTH DATE				CN. BIRTH DATE				CN. BIRTH DATE			
CO. BIRTH DATE				CO. BIRTH DATE				CO. BIRTH DATE			
CP. BIRTH DATE				CP. BIRTH DATE				CP. BIRTH DATE			
CQ. BIRTH DATE				CQ. BIRTH DATE				CQ. BIRTH DATE			
CR. BIRTH DATE				CR. BIRTH DATE				CR. BIRTH DATE			
CS. BIRTH DATE				CS. BIRTH DATE				CS. BIRTH DATE			
CT. BIRTH DATE				CT. BIRTH DATE				CT. BIRTH DATE			
CU. BIRTH DATE				CU. BIRTH DATE				CU. BIRTH DATE			
CV. BIRTH DATE				CV. BIRTH DATE				CV. BIRTH DATE			
CW. BIRTH DATE				CW. BIRTH DATE				CW. BIRTH DATE			
CX. BIRTH DATE				CX. BIRTH DATE				CX. BIRTH DATE			
CY. BIRTH DATE				CY. BIRTH DATE				CY. BIRTH DATE			
CZ. BIRTH DATE				CZ. BIRTH DATE				CZ. BIRTH DATE			
CA. BIRTH DATE				CA. BIRTH DATE				CA. BIRTH DATE			
CB. BIRTH DATE				CB. BIRTH DATE				CB. BIRTH DATE			
CC. BIRTH DATE				CC. BIRTH DATE				CC. BIRTH DATE			
CD. BIRTH DATE				CD. BIRTH DATE				CD. BIRTH DATE			
CE. BIRTH DATE				CE. BIRTH DATE				CE. BIRTH DATE			
CE. BIRTH DATE				CE. BIRTH DATE				CE. BIRTH DATE			
CF. BIRTH DATE				CF. BIRTH DATE				CF. BIRTH DATE			
CG. BIRTH DATE				CG. BIRTH DATE				CG. BIRTH DATE			
CH. BIRTH DATE				CH. BIRTH DATE				CH. BIRTH DATE			
CI. BIRTH DATE				CI. BIRTH DATE				CI. BIRTH DATE			
CJ. BIRTH DATE				CJ. BIRTH DATE				CJ. BIRTH DATE			
CK. BIRTH DATE				CK. BIRTH DATE				CK. BIRTH DATE			
CL. BIRTH DATE				CL. BIRTH DATE				CL. BIRTH DATE			
CM. BIRTH DATE				CM. BIRTH DATE				CM. BIRTH DATE			
CN. BIRTH DATE				CN. BIRTH DATE				CN. BIRTH DATE			
CO. BIRTH DATE				CO. BIRTH DATE				CO. BIRTH DATE			
CP. BIRTH DATE				CP. BIRTH DATE				CP. BIRTH DATE			
CQ. BIRTH DATE				CQ. BIRTH DATE				CQ. BIRTH DATE			
CR. BIRTH DATE				CR. BIRTH DATE				CR. BIRTH DATE			
CS. BIRTH DATE				CS. BIRTH DATE				CS. BIRTH DATE			
CT. BIRTH DATE				CT. BIRTH DATE				CT. BIRTH DATE			
CU. BIRTH DATE				CU. BIRTH DATE				CU. BIRTH DATE			

FIRST SALES	DATE	HOUSE	PLACE
PRESENTATION	24 Feb 74	1400	Broad Ripple RS
RESULTS (SELLING POINTS STRENGTHEN, OBJECTIONS TO BE OVERCOME, ETC.)			
<p>Wants to enlist in WAC ASAP. Living with mother. Divorced last week from a Charles P. Martin, who was chasing other women. Wants to get away from it all. Intends to enlist with her maiden name of Settles. States no problem with police. Studied stenography in H.S. but never used it.</p>			
SUBSEQUENT PRESENTATIONS			
PLACE	DATE	HOUSE	SALES TL
REMARKS Sent to AFEE for Pre-Qual on 25 Feb 74			
RECHIEVER OF CREDIT:			
RELATIVES	FULL NAME	ADDRESS	
MOTHER	Mary Jane Settles	5216 N Boy Scout Road Lawrence, Ind	
FATHER	Deceased		
PERSON	Divorced		
NAME OF RECHIEVER PREPARING CARD		STATION	SOURCE OF LEAD
SFC GEORGE B. HARRIS		Broad Ripple RS	DATE
			24 Feb 74
		<input checked="" type="checkbox"/> BASIC COUNTRY <input type="checkbox"/> OTHER (SPECIFY) Telephone Call	

USE ANSWER CARD A			
ANSWER SHEET - TEST SUPPLEMENT 210-1			
QUALIFICATION CATEGORIES	FURTHER DOCUMENTATION REQUIRED	WAIVABLE DISQUALIFICATION	NOT ELIGIBLE AT THE PRESENT TIME
Name	1	X	X
Age	2	5	4
Citizenship	5	X	6
Social Security Account Number	7	X	X
Moral Qualification	8	9	10
Test Scores	11	12	13
Physical Qualifications	14	15	16
Civilian Education	17	18	19
Marital/Dependency Status	20	21	22
Draft Status	23	X	24
Reserve Status	25	X	26
Prior Military Service	27	28	29
References (Recommendations)	30	X	X

Indicate the grade in which enlistment/reenlistment is authorized assuming the prospect to be fully qualified:

E1	51
E2	52
E3	53
E4	54
E5	55
E6	56
E7	51
E8	52
E9	57
Grade Determination	57

VALID DOCUMENTATION FURNISHED BY PROSPECT

BIRTH CERTIFICATE

Name: Susan Phyllis Settles

Birthdate: Feb 4, 1954

Place of Birth: Kansas City, Missouri

DRIVER'S LICENSE

Issued: Jun 73

Name: Susan S. Martin

Birthdate: Feb 4, 1954

VALID DOCUMENTATION DELIVERED BY PROSPECT

DRIVER'S LICENSE

Issued: Mar '83

Name: Tony L. Henderson

Birthdate: Mar 19, 1938

1. LAST NAME		FIRST NAME		MIDDLE NAME		3 ADDRESS		4 CITY		5 STATE		6 ZIP		7 TELEPHONE	
BIRTH DATE		PLACE		COUNTRY		8 DESCRIPTION		9 RACE		10 SEX		11 NO DEP			
12 SOCIAL SECURITY		13 CLEAR RECORDS		14 TOWNSHIP		15 RACE		16 SEX		17 NO DEP		18 NO DEP			
19 MARITAL STATUS		20 SEX		21 RACE		22 RACE		23 SEX		24 NO DEP		25 NO DEP			
26 SOCIAL SECURITY		27 CLEAR RECORDS		28 TOWNSHIP		29 RACE		30 SEX		31 NO DEP		32 NO DEP			
33 MARITAL STATUS		34 SEX		35 RACE		36 RACE		37 SEX		38 NO DEP		39 NO DEP			
40 SOCIAL SECURITY		41 CLEAR RECORDS		42 TOWNSHIP		43 RACE		44 SEX		45 NO DEP		46 NO DEP			
47 MARITAL STATUS		48 SEX		49 RACE		50 RACE		51 SEX		52 NO DEP		53 NO DEP			
54 SOCIAL SECURITY		55 CLEAR RECORDS		56 TOWNSHIP		57 RACE		58 SEX		59 NO DEP		60 NO DEP			
61 MARITAL STATUS		62 SEX		63 RACE		64 RACE		65 SEX		66 NO DEP		67 NO DEP			
68 SOCIAL SECURITY		69 CLEAR RECORDS		70 TOWNSHIP		71 RACE		72 SEX		73 NO DEP		74 NO DEP			
75 MARITAL STATUS		76 SEX		77 RACE		78 RACE		79 SEX		80 NO DEP		81 NO DEP			
82 SOCIAL SECURITY		83 CLEAR RECORDS		84 TOWNSHIP		85 RACE		86 SEX		87 NO DEP		88 NO DEP			
89 MARITAL STATUS		90 SEX		91 RACE		92 RACE		93 SEX		94 NO DEP		95 NO DEP			
96 SOCIAL SECURITY		97 CLEAR RECORDS		98 TOWNSHIP		99 RACE		100 SEX		101 NO DEP		102 NO DEP			
103 MARITAL STATUS		104 SEX		105 RACE		106 RACE		107 SEX		108 NO DEP		109 NO DEP			
110 SOCIAL SECURITY		111 CLEAR RECORDS		112 TOWNSHIP		113 RACE		114 SEX		115 NO DEP		116 NO DEP			
117 MARITAL STATUS		118 SEX		119 RACE		120 RACE		121 SEX		122 NO DEP		123 NO DEP			
124 SOCIAL SECURITY		125 CLEAR RECORDS		126 TOWNSHIP		127 RACE		128 SEX		129 NO DEP		130 NO DEP			
131 MARITAL STATUS		132 SEX		133 RACE		134 RACE		135 SEX		136 NO DEP		137 NO DEP			
138 SOCIAL SECURITY		139 CLEAR RECORDS		140 TOWNSHIP		141 RACE		142 SEX		143 NO DEP		144 NO DEP			
145 MARITAL STATUS		146 SEX		147 RACE		148 RACE		149 SEX		150 NO DEP		151 NO DEP			
152 SOCIAL SECURITY		153 CLEAR RECORDS		154 TOWNSHIP		155 RACE		156 SEX		157 NO DEP		158 NO DEP			
159 MARITAL STATUS		160 SEX		161 RACE		162 RACE		163 SEX		164 NO DEP		165 NO DEP			
166 SOCIAL SECURITY		167 CLEAR RECORDS		168 TOWNSHIP		169 RACE		170 SEX		171 NO DEP		172 NO DEP			
173 MARITAL STATUS		174 SEX		175 RACE		176 RACE		177 SEX		178 NO DEP		179 NO DEP			
180 SOCIAL SECURITY		181 CLEAR RECORDS		182 TOWNSHIP		183 RACE		184 SEX		185 NO DEP		186 NO DEP			
187 MARITAL STATUS		188 SEX		189 RACE		190 RACE		191 SEX		192 NO DEP		193 NO DEP			
194 SOCIAL SECURITY		195 CLEAR RECORDS		196 TOWNSHIP		197 RACE		198 SEX		199 NO DEP		200 NO DEP			
201 MARITAL STATUS		202 SEX		203 RACE		204 RACE		205 SEX		206 NO DEP		207 NO DEP			
208 SOCIAL SECURITY		209 CLEAR RECORDS		210 TOWNSHIP		211 RACE		212 SEX		213 NO DEP		214 NO DEP			
215 MARITAL STATUS		216 SEX		217 RACE		218 RACE		219 SEX		220 NO DEP		221 NO DEP			
222 SOCIAL SECURITY		223 CLEAR RECORDS		224 TOWNSHIP		225 RACE		226 SEX		227 NO DEP		228 NO DEP			
229 MARITAL STATUS		230 SEX		231 RACE		232 RACE		233 SEX		234 NO DEP		235 NO DEP			
236 SOCIAL SECURITY		237 CLEAR RECORDS		238 TOWNSHIP		239 RACE		240 SEX		241 NO DEP		242 NO DEP			
243 MARITAL STATUS		244 SEX		245 RACE		246 RACE		247 SEX		248 NO DEP		249 NO DEP			
250 SOCIAL SECURITY		251 CLEAR RECORDS		252 TOWNSHIP		253 RACE		254 SEX		255 NO DEP		256 NO DEP			
2															

PRINT LABELS	DATE	HOUR	PLACE
PRESENTATION	21 Feb 74	0800	Broad Ripple RS
STEVIE TITELLING POSTAL SERVICE SUBJECTS TO BE OVERSEEN BY:			
Had been working as construction worker but out of steady work for 9 months. Divorced 3 months ago must support 2 children only. Wife has custody. Has lived in Indianapolis since discharge. Wants to reup immediately. To return on 27 Feb '74 with 714 and 1811.			
UNIDENTIFIED PRESENTATIONS			
PLACE	NAME	NUMBER	DATE
REMARKS: Failed to return on 27 Feb. Couldn't reach him by phone.			
RECEIVED BY CREDIT:			
AGENTS:	FILE NAME	ADDRESS	
MOTHER	Joanne P. Henderson	1217 S. Iowa St. INDIANA CITY, IN	
FATHER	Michael S. Henderson	Same	
SPONSOR	Unrecorded		
NAME OF RECORDED PERSONNEL CARD		STATION	RECEIVED BY
Mt. George B. Harris	Broad Ripple RS	SAK COMB	23 Feb 74

TODAY IS 1 MARCH 1974

ALL DOCUMENTATION FURNISHED BY PROSPECT
 BIRTH CERTIFICATE
 Name: Susan Phyllis Settles
 Birthdate: Feb 4, 1954
 Place of Birth: Kansas City, Missouri
 DRIVER'S LICENSE
 Issued: Jun 73
 Name: Susan S. Martin
 Birthdate: Feb 4, 1954

INCL 7

VALID DOCUMENTATION DELIVERED BY PROSPECT
 DRIVER'S LICENSE
 Issued: Mar 73
 Name: Tony L. Henderson
 Birthdate: Mar 13, 1938

TEST SUPPLEMENT 210-2A

1. NAME (Last, First, Middle Initial) HENDERSON, Tony Thomas		2. ADDRESS Army RA Unasgd		3. TELEPHONE CPL		4. SOCIAL SECURITY NUMBER 284 42 3989	
5. BIRTH DATA DATE: 13 Mar 38 PLACE: Cedar Rapids, Iowa		6. HEIGHT 5' 9 1/2"		7. WEIGHT 140		8. HAIR Brown	
9. MARITAL STATUS SINGLE		10. SELECTIVE SERVICE 3559811407		11. EDUCATION TV-A		12. QUALIFIED FOR ENLISTMENT YES	
13. OCCUPATIONAL STATUS EMPLOYED		14. EMPLOYER Hunting		15. SERVICE JOB INTEREST Airborne		16. MAIN SELLING POINTS Job security retirement	
17. INITIAL PERSON TO PERSON CONTACT YES		18. RECRUITING STATION INDIANAPOLIS		19. OTHER SPECIFIC None		20. TEMP OBJECT None	
21. TEST DATA EST TEST: 51 APPT APNT: 54 EFFICUS: 57		22. DATE 13 Feb 74		23. GRADE GR30		24. PLACE Broad Ripple, IN	
25. RESULTS/SELLING POINTS STRESSED Results: 51, 54, 57 Selling Points: 51, 54, 57		26. DATE 13 Feb 74		27. GRADE GR30		28. PLACE Broad Ripple, IN	
29. RESULTS/SELLING POINTS STRESSED Results: 51, 54, 57 Selling Points: 51, 54, 57		30. DATE 13 Feb 74		31. GRADE GR30		32. PLACE Broad Ripple, IN	

1. NAME (Last, First, Middle Initial) HENDERSON, Tony Thomas		2. ADDRESS 4722 N. Keystone Indianapolis, Ind		3. TELEPHONE 243-7684	
4. BIRTH DATA DATE: 13 Mar 38 PLACE: Cedar Rapids, Iowa		5. HEIGHT 5' 9 1/2"		6. WEIGHT 140	
7. MARITAL STATUS SINGLE		8. SELECTIVE SERVICE 3559811407		9. EDUCATION TV-A	
10. OCCUPATIONAL STATUS EMPLOYED		11. EMPLOYER Hunting		12. SERVICE JOB INTEREST Airborne	
13. INITIAL PERSON TO PERSON CONTACT YES		14. RECRUITING STATION INDIANAPOLIS		15. OTHER SPECIFIC None	
16. TEST DATA EST TEST: 51 APPT APNT: 54 EFFICUS: 57		17. DATE 13 Feb 74		18. GRADE GR30	
19. RESULTS/SELLING POINTS STRESSED Results: 51, 54, 57 Selling Points: 51, 54, 57		20. DATE 13 Feb 74		21. GRADE GR30	
22. RESULTS/SELLING POINTS STRESSED Results: 51, 54, 57 Selling Points: 51, 54, 57		23. DATE 13 Feb 74		24. GRADE GR30	

1. NAME (Last, First, Middle Initial) HENDERSON, Tony Thomas		2. ADDRESS 4722 N. Keystone Indianapolis, Ind		3. TELEPHONE 243-7684	
4. BIRTH DATA DATE: 13 Mar 38 PLACE: Cedar Rapids, Iowa		5. HEIGHT 5' 9 1/2"		6. WEIGHT 140	
7. MARITAL STATUS SINGLE		8. SELECTIVE SERVICE 3559811407		9. EDUCATION TV-A	
10. OCCUPATIONAL STATUS EMPLOYED		11. EMPLOYER Hunting		12. SERVICE JOB INTEREST Airborne	
13. INITIAL PERSON TO PERSON CONTACT YES		14. RECRUITING STATION INDIANAPOLIS		15. OTHER SPECIFIC None	
16. TEST DATA EST TEST: 51 APPT APNT: 54 EFFICUS: 57		17. DATE 13 Feb 74		18. GRADE GR30	
19. RESULTS/SELLING POINTS STRESSED Results: 51, 54, 57 Selling Points: 51, 54, 57		20. DATE 13 Feb 74		21. GRADE GR30	
22. RESULTS/SELLING POINTS STRESSED Results: 51, 54, 57 Selling Points: 51, 54, 57		23. DATE 13 Feb 74		24. GRADE GR30	

ANSWER SHEET - TEST SUPPLEMENT 210-2			
QUALIFICATION CATEGORIES	FURTHER DOCUMENTATION REQUIRED	NAIABLE DISQUALIFICATION	NOT ELIGIBLE AT THE PRESENT TIME
Name	58	X	X
Age	59	X	X
Citizenship	42	X	X
Social Security Account Number	44	X	X
Moral Qualification	45	X	X
Test Scores	48	X	X
Physical Qualifications	51	X	X
Civilian Education	54	X	X
Marital/Dependency Status	57	X	X
Draft Status	60	X	X
Reserve Status	62	X	X
Prior Military Service	64	X	X
References (Recommendations)	67	X	X
Indicate the grade in which enlistment/reenlistment is authorized assuming the prospect to be fully qualified			
E1	68		
E2	69		
E3	70		
E4	71		
E5	72		
E6	73		
E7	68		
E8	69		
E9	74		
Grade Determination	74		

TEST SUPPLEMENT 210-3A

TODAY IS 1 MARCH 1974

1. NAME (Last, First, Middle, Initial) JONES, HARVEY PRICE		2. SOCIAL SECURITY NUMBER 812 31 3455	
3. GRADE SSG		4. DATE OF BIRTH 1 June 1973	
5. ADDRESS (Street, City, State, Zip) Indianapolis, Marion County, Indiana 46233		6. DATE OF ENTRY INTO SERVICE 1 July 1973	
7. TYPE OF SERVICE Discharged		8. DATE OF DISCHARGE 1 Jan 1974	
9. REASON FOR DISCHARGE AR 635-200 SPN 201		10. TYPE OF DISCHARGE Honorable	
11. PLACE OF DISCHARGE Fort Benjamin Harrison, Indiana 46216		12. DATE OF DISCHARGE 1 Jan 1974	
13. TYPE OF SERVICE One		14. DATE OF ENTRY INTO SERVICE 1 July 1973	
15. ADDRESS (Street, City, State, Zip) 4112 North Franklin Road, Indianapolis, Indiana 46226		16. DATE OF ENTRY INTO SERVICE 1 July 1973	
17. TYPE OF SERVICE 71140 Admin Spec		18. DATE OF ENTRY INTO SERVICE 1 July 1973	
19. TYPE OF SERVICE 169268 Manager, Office		20. DATE OF ENTRY INTO SERVICE 1 July 1973	
21. TYPE OF SERVICE National Defense Service Medal, Vietnam Campaign Medal, Good Conduct Medal (2nd Award), Army Commendation Medal		22. DATE OF ENTRY INTO SERVICE 1 July 1973	
23. TYPE OF SERVICE One		24. DATE OF ENTRY INTO SERVICE 1 July 1973	
25. TYPE OF SERVICE One		26. DATE OF ENTRY INTO SERVICE 1 July 1973	
27. TYPE OF SERVICE One		28. DATE OF ENTRY INTO SERVICE 1 July 1973	
29. TYPE OF SERVICE One		30. DATE OF ENTRY INTO SERVICE 1 July 1973	
31. TYPE OF SERVICE One		32. DATE OF ENTRY INTO SERVICE 1 July 1973	
33. TYPE OF SERVICE One		34. DATE OF ENTRY INTO SERVICE 1 July 1973	
35. TYPE OF SERVICE One		36. DATE OF ENTRY INTO SERVICE 1 July 1973	
37. TYPE OF SERVICE One		38. DATE OF ENTRY INTO SERVICE 1 July 1973	
39. TYPE OF SERVICE One		40. DATE OF ENTRY INTO SERVICE 1 July 1973	
41. TYPE OF SERVICE One		42. DATE OF ENTRY INTO SERVICE 1 July 1973	
43. TYPE OF SERVICE One		44. DATE OF ENTRY INTO SERVICE 1 July 1973	
45. TYPE OF SERVICE One		46. DATE OF ENTRY INTO SERVICE 1 July 1973	
47. TYPE OF SERVICE One		48. DATE OF ENTRY INTO SERVICE 1 July 1973	
49. TYPE OF SERVICE One		50. DATE OF ENTRY INTO SERVICE 1 July 1973	
51. TYPE OF SERVICE One		52. DATE OF ENTRY INTO SERVICE 1 July 1973	
53. TYPE OF SERVICE One		54. DATE OF ENTRY INTO SERVICE 1 July 1973	
55. TYPE OF SERVICE One		56. DATE OF ENTRY INTO SERVICE 1 July 1973	
57. TYPE OF SERVICE One		58. DATE OF ENTRY INTO SERVICE 1 July 1973	
59. TYPE OF SERVICE One		60. DATE OF ENTRY INTO SERVICE 1 July 1973	
61. TYPE OF SERVICE One		62. DATE OF ENTRY INTO SERVICE 1 July 1973	
63. TYPE OF SERVICE One		64. DATE OF ENTRY INTO SERVICE 1 July 1973	
65. TYPE OF SERVICE One		66. DATE OF ENTRY INTO SERVICE 1 July 1973	
67. TYPE OF SERVICE One		68. DATE OF ENTRY INTO SERVICE 1 July 1973	
69. TYPE OF SERVICE One		70. DATE OF ENTRY INTO SERVICE 1 July 1973	
71. TYPE OF SERVICE One		72. DATE OF ENTRY INTO SERVICE 1 July 1973	
73. TYPE OF SERVICE One		74. DATE OF ENTRY INTO SERVICE 1 July 1973	
75. TYPE OF SERVICE One		76. DATE OF ENTRY INTO SERVICE 1 July 1973	
77. TYPE OF SERVICE One		78. DATE OF ENTRY INTO SERVICE 1 July 1973	
79. TYPE OF SERVICE One		80. DATE OF ENTRY INTO SERVICE 1 July 1973	
81. TYPE OF SERVICE One		82. DATE OF ENTRY INTO SERVICE 1 July 1973	
83. TYPE OF SERVICE One		84. DATE OF ENTRY INTO SERVICE 1 July 1973	
85. TYPE OF SERVICE One		86. DATE OF ENTRY INTO SERVICE 1 July 1973	
87. TYPE OF SERVICE One		88. DATE OF ENTRY INTO SERVICE 1 July 1973	
89. TYPE OF SERVICE One		90. DATE OF ENTRY INTO SERVICE 1 July 1973	
91. TYPE OF SERVICE One		92. DATE OF ENTRY INTO SERVICE 1 July 1973	
93. TYPE OF SERVICE One		94. DATE OF ENTRY INTO SERVICE 1 July 1973	
95. TYPE OF SERVICE One		96. DATE OF ENTRY INTO SERVICE 1 July 1973	
97. TYPE OF SERVICE One		98. DATE OF ENTRY INTO SERVICE 1 July 1973	
99. TYPE OF SERVICE One		100. DATE OF ENTRY INTO SERVICE 1 July 1973	

VALID DOCUMENTATION DELIVERED BY PROSPECT

DRIVER'S LICENSE

Issued: 14 January 1974

Name: Harvey Price Jones

Birthdate: 1 July 1973

USE ANSWER CARD B

ANSWER SHEET - TEST SUPPLEMENT 210-3

QUALIFICATION CATEGORIES	FURTHER DOCUMENTATION REQUIRED	WAIVABLE DISQUALIFICATION	NOT ELIGIBLE AT THE PRESENT TIME
Name	1	X	X
Age	2	3	4
Citizenship	5	X	6
Social Security Account Number	7	X	8
Moral Qualification	9	10	11
Test Scores	12	13	14
Physical Qualifications	15	16	17
Civilian Education	18	19	20
Marital/Dependency Status	21	22	23
Draft Status	24	X	25
Reserve Status	26	X	27
Prior Military Service	28	29	30
References (Recommendations)	31	X	32

Indicate the grade in which enlistment/reenlistment is authorized assuming the prospect to be fully qualified

1.1	31
1.2	32
1.3	33
1.4	34
1.5	35
1.6	36
1.7	37
1.8	38
1.9	39
Grade Determination	37

1. LAST NAME JONES, Harvey Price		2. ADDRESS 4112 N Franklin Rd, Marion County, Indianapolis, IN 46226	
3. BIRTH DATA DATE: 1 July 73, PLACE: Indianapolis, IN		4. SOCIAL SECURITY NUMBER 812-31-3455	
5. MARITAL STATUS SINGLE		6. EDUCATION HS	
7. OCCUPATIONAL STATUS EMPLOYED		8. TYPE OF SERVICE RE-ENTRY	
9. PRIOR ACTIVE SERVICE YES		10. RESERVE STATUS READY	
11. MONTHS 81		12. GRADE SSG	
13. DATE SEPARATED 1 Jan 74		14. COMPONENT USAR	
15. SERVICE NUMBER 812-31-3455		16. UNIT 425	
17. INITIAL PERSON TO PERSON CONTACT RECRUITING STATION		18. OTHER (SPECIFY)	
19. TESTS DATE: 26 Feb 74, SCORES: 111, 122, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150		20. REASON FOR REJECTION None	
21. ENLISTED DEF DATE 1 July 73		22. ENLISTED RA DATE 1 July 73	

1. FIRST NAME Harvey		2. LAST NAME Jones	
3. DATE 27 Feb 74		4. HOUR 0930	
5. PLACE Recruiting Main Station, Indianapolis, IN		6. RESULTS (SELLING POINTS STRESSED OBJECTIONS TO BE OVERCOME ETC.)	
<p>Harvey stated that he has not been able to find a job that has suited him since he was discharged. He said his wife is pregnant again so he would like to reenlist as soon as possible.</p>			
7. DATE 27 Feb 74		8. HOUR 0930	
9. PLACE Recruiting Main Station, Indianapolis, IN		10. RESULTS (SELLING POINTS STRESSED OBJECTIONS TO BE OVERCOME ETC.)	
<p>None</p>			
11. DATE 27 Feb 74		12. HOUR 0930	
13. PLACE Recruiting Main Station, Indianapolis, IN		14. RESULTS (SELLING POINTS STRESSED OBJECTIONS TO BE OVERCOME ETC.)	
<p>None</p>			
15. DATE 27 Feb 74		16. HOUR 0930	
17. PLACE Recruiting Main Station, Indianapolis, IN		18. RESULTS (SELLING POINTS STRESSED OBJECTIONS TO BE OVERCOME ETC.)	
<p>None</p>			
19. DATE 27 Feb 74		20. HOUR 0930	
21. PLACE Recruiting Main Station, Indianapolis, IN		22. RESULTS (SELLING POINTS STRESSED OBJECTIONS TO BE OVERCOME ETC.)	
<p>None</p>			

TODAY IS 1 MARCH 1974

TEST SUPPLEMENT 210-4A

1. LAST NAME MILLS, Ralph Duane		2. ADDRESS 9110 Conried Dr, Marion, Indianapolis, IN 46250		3. TELEPHONE 546-2207	
4. BIRTH DATA DATE: 21 Mar 50 PLACE: NY, N.Y.		5. DESCRIPTION SEX: M, RACE: W, HEIGHT: 5'10", WEIGHT: 187, EYES: BRN, HAIR: BRN		6. AGE 24	
7. MARITAL STATUS MARRIED: YES, DIVORCED: NO, SEPARATED: NO, SINGLE: YES		8. SELECTIVE SERVICE NUMBER: 12345678, CLASSIFICATION: 1-A, GRADE: 1-A		9. EDUCATION YEARS: 10, GRADUATED: YES, SCHOOL: [blank]	
10. EMPLOYMENT EMPLOYED: YES, UNEMPLOYED: NO, OCCUPATION: Custodian, John Constr Co, Indpls, IN		11. LISTED FOR FOLLOW-UP DATE: 2 Mar 74, HOUR: 1100		12. QUALIFIED FOR ENLISTMENT IN GRADE YES: YES, NO: NO	
13. PRIOR ACTIVE SERVICE YES: NO, GRADE: [blank], DATE DEPARTED: [blank]		14. RESERVE STATUS READY: YES, STANDBY: NO, DEP: YES		15. SERVICE JOB INTEREST YES: YES, NO: NO	
16. DATES SALES LETTERS SENT TO: [blank]		17. SERVICE JOB INTEREST YES: YES, NO: NO		18. BAIN SELLING POINTS Model Railroad	
19. INITIAL PERSON TO PERSON CONTACT RECRUITING STATION: [blank], OTHER SPECIFY: [blank]					
20. TESTS EST WEST: 24 Feb 74, 29, DATE: 24 Feb 74, PPM: 111122, REASON FOR REJECTION: [blank]					
21. APPT APPLY DATE: 24 Feb 74, COA: COM, 105, 95, 85, 75, 65, 55, 45, 35, 25, 15, 5, 0					
22. ENLISTED DEP DATE PERIOD: [blank], OPTION: [blank]					
23. TO ENLIST IN GRADE PERIOD: [blank], OPTION: [blank]					
24. ENLISTED AS DATE PERIOD: [blank], OPTION: [blank]					

USAREC FORM 100
28 FEB 68

RECRUITING PROSPECT CARD
USAREC REG 01-20

FIRST SALES DATE: 24 Feb 74, HOUR: 1030, PLACE: Broad Ripple Recruiting Station	
PRESENTATION RESULTS: [blank]	
REMARKS Sent to APBS for Pre-Qual 24 Feb 74	
SUBSEQUENT PRESENTATIONS PLACE: [blank], DATE: [blank], HOUR: [blank], RESULTS: [blank]	
RECRUITER OF CREDIT RELATIVE: [blank], FULL NAME: [blank], ADDRESS: [blank]	
OTHER: Mills, Rebecca Sue, 1181 N. College, Indpls, IN	
FATHER: Mills, Ralph Edward, Same as above	
GRADE: [blank]	
WANT OF RECRUITER PREPARING CARD STATION: Broad Ripple, IN, SUBJECT OF CARD: [blank], DATE: 24 Feb 74	

USE ANSWER CARD B

* ANSWER SHEET FOR TEST SUPPLEMENT 210-4

QUALIFICATION CATEGORIES	FURTHER DOCUMENTATION REQUIRED	WATVABLE DISQUALIFICATION	NOT ELIGIBLE AT THE PRESENT TIME
Name	18	X	X
Age	39	40	41
Citizenship	42	X	43
Social Security Account Number	44	X	X
Moral Qualification	45	46	47
Test Scores	48	49	50
Physical Qualifications	51	52	53
Civilian Education	54	55	56
Marital/Dependency Status	57	58	59
Draft Status	60	X	61
Reserve Status	62	X	63
Prior Military Service	64	65	66
References (Recommendations)	67	X	X

Indicate the grade in which enlistment/reenlistment is authorized assuming the prospect to be fully qualified:

E1	68
E2	69
E3	70
E4	71
E5	72
E6	73
E7	68
E8	69
E9	74
Grade Determination	74

1. LAST NAME MILLS, Ralph Duane		2. ADDRESS 9110 Conried Dr, Marion, Indianapolis, IN 46250		3. TELEPHONE 546-2207	
4. BIRTH DATA DATE: 21 Mar 50 PLACE: NY, N.Y.		5. DESCRIPTION SEX: M, RACE: W, HEIGHT: 5'10", WEIGHT: 187, EYES: BRN, HAIR: BRN		6. AGE 24	
7. MARITAL STATUS MARRIED: YES, DIVORCED: NO, SEPARATED: NO, SINGLE: YES		8. SELECTIVE SERVICE NUMBER: 12345678, CLASSIFICATION: 1-A, GRADE: 1-A		9. EDUCATION YEARS: 10, GRADUATED: YES, SCHOOL: [blank]	
10. EMPLOYMENT EMPLOYED: YES, UNEMPLOYED: NO, OCCUPATION: Custodian, John Constr Co, Indpls, IN		11. LISTED FOR FOLLOW-UP DATE: 2 Mar 74, HOUR: 1100		12. QUALIFIED FOR ENLISTMENT IN GRADE YES: YES, NO: NO	
13. PRIOR ACTIVE SERVICE YES: NO, GRADE: [blank], DATE DEPARTED: [blank]		14. RESERVE STATUS READY: YES, STANDBY: NO, DEP: YES		15. SERVICE JOB INTEREST YES: YES, NO: NO	
16. DATES SALES LETTERS SENT TO: [blank]		17. SERVICE JOB INTEREST YES: YES, NO: NO		18. BAIN SELLING POINTS Model Railroad	
19. INITIAL PERSON TO PERSON CONTACT RECRUITING STATION: [blank], OTHER SPECIFY: [blank]					
20. TESTS EST WEST: 24 Feb 74, 29, DATE: 24 Feb 74, PPM: 111122, REASON FOR REJECTION: [blank]					
21. APPT APPLY DATE: 24 Feb 74, COA: COM, 105, 95, 85, 75, 65, 55, 45, 35, 25, 15, 5, 0					
22. ENLISTED DEP DATE PERIOD: [blank], OPTION: [blank]					
23. TO ENLIST IN GRADE PERIOD: [blank], OPTION: [blank]					
24. ENLISTED AS DATE PERIOD: [blank], OPTION: [blank]					

USAREC FORM 100
28 FEB 68

RECRUITING PROSPECT CARD
USAREC REG 01-20

FIRST SALES DATE: 24 Feb 74, HOUR: 1030, PLACE: Broad Ripple Recruiting Station	
PRESENTATION RESULTS: [blank]	
REMARKS Sent to APBS for Pre-Qual 24 Feb 74	
SUBSEQUENT PRESENTATIONS PLACE: [blank], DATE: [blank], HOUR: [blank], RESULTS: [blank]	
RECRUITER OF CREDIT RELATIVE: [blank], FULL NAME: [blank], ADDRESS: [blank]	
OTHER: Mills, Rebecca Sue, 1181 N. College, Indpls, IN	
FATHER: Mills, Ralph Edward, Same as above	
GRADE: [blank]	
WANT OF RECRUITER PREPARING CARD STATION: Broad Ripple, IN, SUBJECT OF CARD: [blank], DATE: 24 Feb 74	

*** ITEM ANALYSIS ***

COURSE - AR&CC
CLASS - 72-01

TEST - QUAL 210
SITUATION - A

CARD NO - 1 OF 3
RAW POSS - 36

QUES.	PUN.	UNPUN.	MISS RATE	QUES.	PUN.	UNPUN.	MISS RATE
1	4*	46	92%***	19		50*	
2	32*	18	36%***	20	28*	22	44%***
3	2	48*	4%	21		50*	
4	1	49*	2%	22		50*	
5		50*		23	2	48*	4%
6		50*		24		50*	
7	49*	1	2%	25		50*	
8	48*	2	4%	26		50*	
9		50*		27		50*	
10		50*		28		50*	
11	1	49*	2%	29		50*	
12	1	49*	2%	30	49*	1	2%
13	26*	24	48%***	31	48*	2	4%
14	5	45*	10%	32		50*	
15		50*		33		50*	
16		50*		34		50*	
17	45*	5	10%	35		50*	
18		50*		36		50*	

ELIGIBILITY 210-1A (SETTLES)
ITEM ANALYSIS - MISS RATES

<u>ITEM #</u>	CYCLE 1			
	<u>CLASS</u> <u>72-01</u>	<u>CLASS</u> <u>72-02</u>	<u>CLASS</u> <u>72-04</u>	<u>CLASS</u> <u>72-05</u>
1	92	32	64	42
2	36	51	36	34
4	2	12	8	10
5	0	14	8	6
8	4	0	4	2
11	2	4	6	0
13	48	53	40	18
14	10	10	12	8
17	10	14	16	18
20	44	26	34	30
23	4	0	2	0
30	2	18	24	0

ELIGIBILITY 210-1A (SETTLES)
ITEM ANALYSIS MISS RATES (BY INSTRUCTOR)

MISS RATES

ITEM NO.	SGT D		SGT E	
	CYCLE 1	CYCLE 2	CYCLE 1	CYCLE 2
1	92*	4	42*	8
2	36*	20*	34*	8
4	2	10	10	2
5	0	2	6	2
13	48*	38*	18*	10
14	10	6	8	4
17	10	12*	18*	8
20	44*	24*	30*	0
30	2	2	0	8

UNITED STATES ARMY ADJUTANT GENERAL SCHOOL



SYSTEMS ENGINEERING OF TRAINING

LOGIC TREE FOR

UTILIZING ITEM ANALYSIS TEST RESULTS (SET 7-1)

JOB: Training Analyst or Instructor

TASK: To utilize an item analysis test result in order to eliminate deficiencies in instructor test design

CUE: Receipt of an item analysis test result printout

CONDITIONS: Access to the appropriate answer/score sheet key, test instrument and lesson plan

Prepared for the Quality Control portion of the CONARC Training Workshop, OCT 5-7, 1971. (To be incorporated into the USAAGS Instructor Training Course)

Comments and/or suggestions should be forwarded to USAAGS, ATTN: Director of Instruction. Telephone - Autovon 699-3648 or 699-3703; civilian (317) 542-3648 or 542-3703

Source

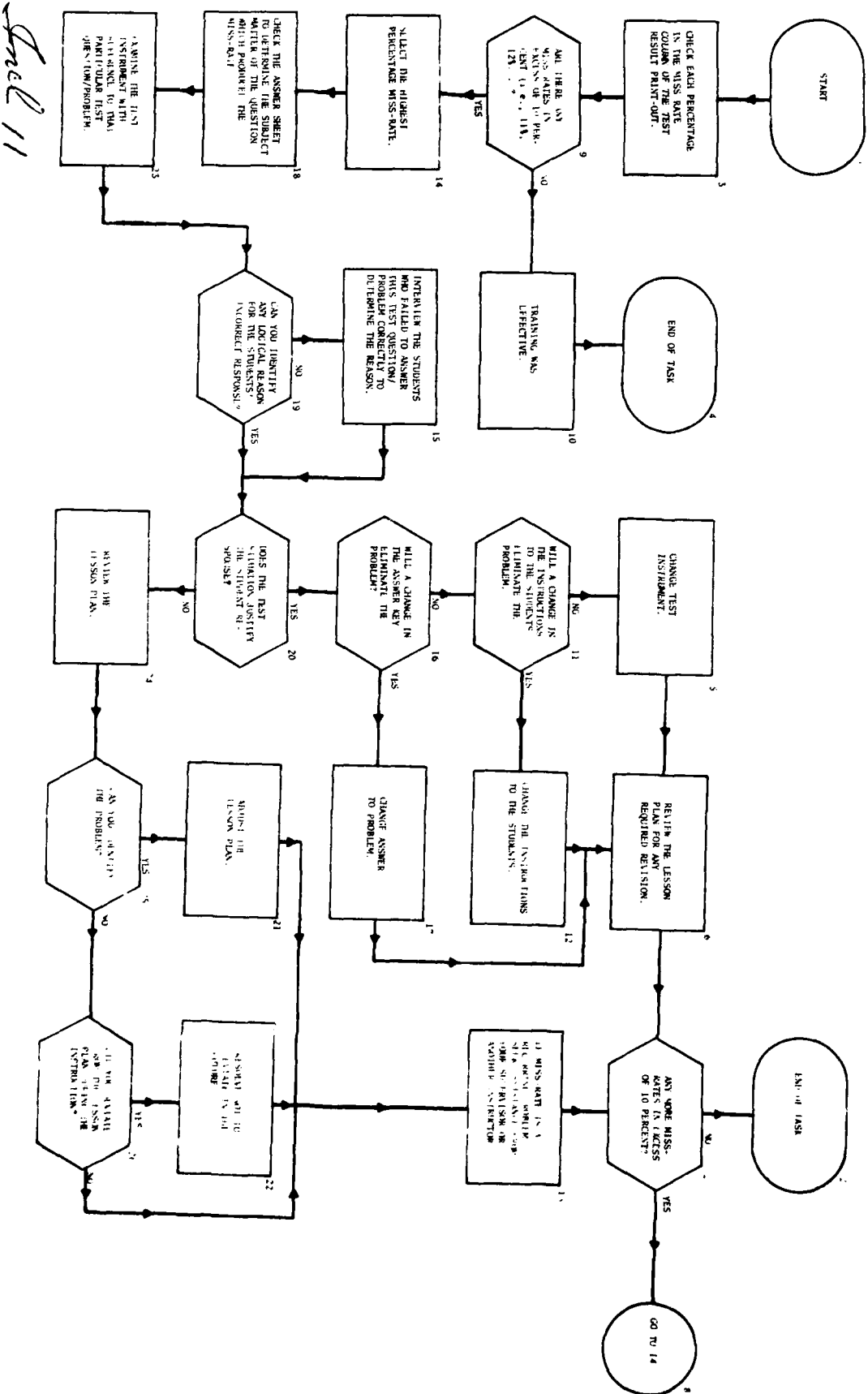
SET 7-1
AUG 71

Con Reg 350-100-1 (SET) dtd 1 Feb 68 w/C1
USAAGS Reg 350-100 (SET) dtd 1 Jul 68 w/C1
USAAGS Reg 350-2 (Resident Student Evaluation Program) dtd 30 Jan 69
USAAGS Local SOP



FORT BENJAMIN HARRISON, INDIANA 46216

Incl 11



CASH MONEY ORDERS
ITEM ANALYSIS MISS RATES

	<u>69-03</u>	<u>69-04</u>	<u>72-01</u>
1. (AMOUNT PAID)	?	?	2
2. (5 M.O. INCLOSED)	0	2	2
3. (INDORSEMENT)	14*	4	0
4. (SSAN)	41*	7	0
5. (PAYEE)	16*	4	0
6. (ID)	10	2	2
7. (APDS)	6	2	--
8. (INITIALS)	6	2	2
9. (ONE M.O. UNPAID)	14*	7	5
<u>MONEY ORDER</u>			
POI HRS	26	26	16

INCL 13 (Incl 12 is omitted, see note page VII-22)

PERFORMANCE TEST FOR 083 SORTER

NAME

STUDENT NUMBER

DATE _____

- | | | | |
|------|---|-----|----|
| 1. | CARDS JOGGLED PRIOR TO BEING PLACED IN HOPPER | YES | NO |
| 2. | PROPER PLACEMENT OF CARDS IN HOPPER. | YES | NO |
| 3. | SUPPRESSION KEYS CHECKED. | YES | NO |
| 4. | SUPPRESSION KEYS PROPERLY SET FOR SELECTION. | YES | NO |
| 5. | SORT TEST SWITCH IN SORT POSITION. | YES | NO |
| 6. | EDIT AND EDIT STOP SWITCHES IN OFF POSITION. | YES | NO |
| 7. | SORT SELECTION SWITCH SET ON-N-FOR NUMERIC SORTING. | YES | NO |
| 8. | SORT SELECTION SWITCH SET ON-N-AND-Z-FOR ALPHA SORTING. | YES | NO |
| *9. | SORT BRUSH SET FOR MINOR SORT (FIRST SORT). | YES | NO |
| *10. | SORT BRUSH SET FOR INTERMEDIATE SORT (SECOND SORT). | YES | NO |
| *11. | SORT BRUSH SET FOR SELECTION SORT (THIRD SORT). | YES | NO |
| *12. | PROPER REMOVAL OF CARDS FROM MACHINE. | YES | NO |
| 13. | SIGHT CHECK PROPERLY PERFORMED. | YES | NO |
| 14. | CARD COUNT MADE AT BEGINNING AND END OF OPERATION. | YES | NO |
| 15. | CARDS IN CORRECT SEQUENCE WHEN RETURNED TO INSTRUCTOR. | YES | NO |

***LAST CHANCE FOR STUDENT TO CORRECT SET-UP ERROR WITHOUT PENALTY**

[illegible]

QUALITY CONTROL WORKSHOP

PROBLEMS IN THE DESIGN OF SUBJECTIVELY SCORED PERFORMANCE TESTS

By ROBERT N. JOHNSON, GS-13
Asst Director of Instruction, USAAGS

Up to this point we have confined our discussion to the design of objectively scored performance tests. By sequencing the presentation in this order we put our best foot forward, but now we are left with what is really our major problem area; that is developing useful feedback on subjectively scored tests, both Product Scored and Observer Scored. For the next 10 minutes I'd like to address this area of consideration. I am sure that many of you can think of tasks which cannot be measured on an objective basis but to be sure that we are all on the same wave length let me outline two tasks as an example of the problem.

For our Product Scored example, let's take a task from the area of written communications. Graduates of the AG Officer Basic Course who go on duty as a Personnel Actions Officer must reply to letters from the next-of-kin who request information concerning the circumstances under which a son or husband was killed in action. Creation of a realistic test instrument is really very simple. Students are given the letter from the next-of-kin together with the actual details of the death. The requirement - to write the reply. The problem is, of course, in how to score and grade the product - the reply written by the students. Some elements of the reply, such as the inclusion of an answer to each specific question asked by the next-of-kin, can be scored objectively on a go/no go basis. Other elements such as the tone of the reply, its organization and conciseness cannot be objectively measured.

For our Observer Scored example, let's take a case of a recruiter in a civilian community who must present a formal presentation to a civic group or to a class of high school students. This is oral communications - speech - pure and simple. Student performance in this case, must be observed by the scorer. Again it is easy to develop the requirement but how do you score the students performance. Many approaches have been tried over the years. The usual approach being to have a "qualified" instructor observe the performance or score the product based on a list of required elements and then assign an alphabetic or adjectival grade. In some cases the elements of task performance are weighted, in others they are not. The obvious purpose of these approaches is to assign a grade

to the individual student. If the observer or scorer is indeed qualified, the student also receive constructive feedback. No attempt is made, however, to examine feedback, for the purpose of improving the instructional system itself, as we discussed earlier today.

The AG School, therefore, decided to approach the problem of evaluation of oral and written communication tasks from a quality control standpoint by trying to isolate student weaknesses which could serve as the basis for modification of curriculum. Our efforts to date have accomplished little more than to more clearly identify the problem through the use of computerized feedback.

Our approach to scoring problem was to limit the subjectivity of the scorer to a three point scale for each supposed element of performance. That is, the scorer rates each element on the score sheet as either weak, OK, or outstanding. The weak and outstanding categories representing the extremes only. That is, these categories were to be used only when the element being rated was clearly weak or clearly outstanding. We took this approach thinking that extremes are more easily and validly identified by instructors.

The grading concept was very simple. If, for instance there are 26 factors which comprise a good speech, the more of these factors you have working for you and the fewer you have working against you, the better the chances are, that the overall effect of the speech will be satisfactory. According we use a minus, zero, plus grading method. Each weak score is a minus one, each outstanding score is a plus one, average scores receive no weight. The total raw therefore is the outstanding scores minus the weak scores. To avoid negative values and to fit the system into our current computer program we designed the following score sheet for evaluating speeches. (Incl 1 - Score Sheet). Note that the observer scores each item independently by circling a number which will later be punched into a punch card.

We applied this method to our first eight classes in the Recruiting course. We did not attempt to establish a P/F point for grading purpose but rather allowed the computer to grade each class separately on a normal curve basis. The range and dispersion of scores was so wide that not a single student failed the test on a norm referenced basis. Our item analysis however did identify what appeared to be problem areas or student weaknesses. For example in seven classes out of eight, gestures, movement and eye contact showed up as problem areas. The question now arose, were the student scores and the item miss rates representative of student performance or were they merely indicators of instructor bias. In an attempt to answer this question we took the 400 speech punch cards from the cycle and organized them by the scorer rather than by class and again ran

them through the computer with rather humorous results. For this run we selected a P/F point based on the normal curve of the entire eight classes. (Incl 2 VG - Speech Scores). This is only a sample of the results of 31 observers. Note that the average raw scores run from a low 32.7 to a high of 53.1. But look closely at how the scores were distributed. Observers 1 and 7 are both tough but Observer 7 never gave anyone over 79%. Observer 6 loves everyone.

In the item analysis our toughest rater found a full 70% of his student weak on eye contact, 55% weak on the summary and 50% weak on gestures and movement. The only item for which he did not dock anyone was sincerity. Apparently everyone was sincere. Our friendliest rater found no one weak in anything. Gave 70% an outstanding on their summary and closing statement and generally handed out outstanding ratings like candy to babies. An average rater found student weakness in only two areas, support of main points and eye contact, about 10 and 20% respectively. He, however, gave away very few outstandings so his results were a little above average.

It became obvious at this point that our results were definitely affected by instructor bias. What did we do? What any good manager does - start a training program! The instructional division chief spend an entire day showing video tapes of actual student performances and critiquing the performance on a item by item basis with the entire group of instructors. At the end of the training all instructors scored a single Video tape with the following very disheartening results. (Incl 3 VG - Speech Scores after Training). The division chief rated this individual as a 26. Note the wide range of scores even after training. The item analysis is even more revealing. Here is an extract from the item analysis on this case. (Incl 4 VG - Single Speech Item Analysis). Note the limited amount of agreement on rating factors even after training. The difference between a weak and OK is admittedly a thin line but how about the differences between a weak and an outstanding as shown in item 4.

Some of you may feel that the basic problem is in the numeric system which we tried and that your alphabetic or adjectival scoring systems are much more effective. I am sure, however, that if you look deeper into your own system you will find the identical problems involved. I've merely tried to present the problem for discussion. Now, what do you think? How do you approach this type problem in your school?

THE FOLLOWING IS A SUMMARY OF THE DISCUSSION WHICH FOLLOWED THE ABOVE PRESENTATION

Inability to design valid tests which measure "soft-skill" tasks requiring subjective evaluation was recognized as a major problem. The type tasks causing the most difficulty appeared to be those which involved communication between two individuals wherein one person is attempting to influence or alter the attitude or behavior of the other. This type task appears frequently in officer classes and in many of the "soft-skill" enlisted classes. Many approaches were discussed from peer ratings and self ratings to possibility of elimination of the requirement for a test in these areas. No solution or agreement was reached.

SCORE SHEET FOR FORMAL PRESENTATION

STUDENT NUMBER	STUDENT NAME	DATE	EVALUATOR	COURSE	TYPE OF SPEECH
A. WEAK B. OK C. OUTSTANDING	Rate the speaker on each item shown below by circling the appropriate number. Identify in the Comments Section any portion of the presentation that significantly adds or detracts from its overall effectiveness.				
A	B	C	I - INTRODUCTION		COMMENTS
1	99	37	1. LEAD IN.		
2	99	38	2. MOTIVATION.		
3	99		3. THEME.		
			II - BODY		
4	99	40	4. SELECTION OF MAIN POINTS.		
5	99	41	5. EMPHASIS ON MAIN POINTS.		
6	99	42	6. SUPPORT OF MAIN POINTS.		
7	99		7. SEQUENCING OF MATERIALS.		
8	99	44	8. TRANSITION BETWEEN POINTS.		
9	99	45	9. SUBJECT KNOWLEDGE.		
			III - VOICE AND SPEECH TECHNIQUE		
10	99		10. EYE CONTACT.		
11	99		11. VOCAL VARIETY.		
12	99		12. VOLUME.		
13	99		13. RATE OF DELIVERY.		
14	99		14. PRONUNCIATION AND GRAMMAR.		
15	99		15. VOCABULARY LEVEL.		
16	99		16. ENUNCIATION.		

Inc 1

			IV - APPEARANCE
17	99		17. BEARING AND POSTURE.
18	99		18. GESTURES AND MOVEMENT.
			V - PERSONALITY
19	99	55	19. SINCERITY.
20	99	56	20. CONFIDENCE.
21	99	57	21. ENTHUSIASM.
22	99	58	22. AUDIENCE RAPPORT.
			VI - CONCLUSION
23	99		23. QUESTIONS.
24	99		24. SUMMARY.
25	99	61	25. CLOSING STATEMENT.
26	99		26. OVERALL TIMING.
27-36	99	63	27. OVERALL IMPRESSION.

FBH FORM 2-13
18 AUG 71

REPLACES FBH FORM 2-13, 18 MAY 71, WHICH IS OBSOLETE.

INCL # 1

2

SPEECH SCORES BY OBSERVER

MAX RAW = 63

<u>OBSERVER</u>	<u>AVERAGE RAW</u>	<u>F</u>	<u>L (70-79)</u>	<u>M (80-89)</u>	<u>U (90-100)</u>
1	38.6	38	25	25	13
2	41.5		70	30	
3	49.1		11	58	32
4	39.6		85	15	
5	37.1	20	60	20	
6	53.1			20	80
7	32.7	20	80		
8	35.5	6	88	6	
9	46.3		11	67	22

Incl #2

SINGLE TECH SCORES (VIDEO TAPE)
AFTER TRAINING

43-44	2	XX	
41-42	0		
39-40	2	XX	
37-38	1	X	AVERAGE 30.9
35-36	5	XXXXX	
33-34	5	XXXXX	
31-32	8	XXXXXXXXX	P/F
29-30	7	XXXXXXXXX	
27-28	2	XX	
25-26	10	XXXXXXXXXXXX	
23-24	1	X	
21-22	1	X	

SINGL 'EECH ITEM ANALYSIS
AFTER TRAINING

	<u>WEAK</u>	<u>OK</u>	<u>OUTSTANDING</u>
1 LEAD-IN	75%	25%	----
2 MOTIVATION	61%	39%	----
4 SEL OF MAIN POINTS	11%	75%	14%
18 GESTURES	50%	50%	----
20 CONFIDENCE	27%	66%	7%
21 ENTHUSIASM	65%	30%	5%

Incl #14

QUALITY CONTROL WORKSHOP

TEST SCORING/GRADING

By MR. JAMES L. SHERRILL/
MAJ WILLIAM V. GREEN

To insure that we are all talking the same language, let me use an example to explain the difference between the terms "scoring" and "grading."

Assume you are an instructor at one of the CONARC Schools. You administer the following test situation to one of your students:

SITUATION

Task: Remove, "bench-time", then replace the right magneto.

You have a performance checklist that indicates the task has 11 scorable items that total 30 raw points on the grading plan. As you observe the student's performance, you notice that he does everything correctly, but has worked on the left magneto.

How will you score the student?

- A. Give full credit.
- B. Take a few points.
- C. Take most of the points.
- D. Fail the student on that test.

Obviously the above choices do not include the correct answer. Since we are concerned with "scoring" and not grading, we should not be evaluating the student's performance. When a test is scored, you should faithfully record the performance exactly as it occurs.

Only after you have determined how the test will be scored do you consider grading or evaluating. Then you must answer questions such as: Do I have a job standard? How perfect must this task be performed? Should we weight any test items?

Items are weighted because the training analyst feels the item is

either critical or should be worth more in determining the overall grade.

We have argued over the question of weighting test items for the last few years. Although there may be valid reasons why task elements should be weighted, we do not weight any item. Let me take a few minutes to discuss our philosophy with you.

Assume we have a test situation with five scorable items and we administer it to ten students. The results of the test are shown below:

TEST RESULTS

Frequency Distribution

<u>Total Raw Scored</u>	<u>Number of Students</u>
30	2
24	3
21	3
15	1
12	1

Raw Possible - 30

ITEM ANALYSIS

<u>Item (Weight)</u>	<u>Percent Miss-Rate</u>
1 (1)	50
2 (2)	20
3 (2)	80
4 (1)	90
5 (3)	0

Item 5 would appear to have the most effect on ranking students; however, this is not true in this case. As shown below, items that

all students passed have no impact on overall class ranking.

ANALYSIS

<u>Item</u>	<u>Percent Correct (P)</u>	<u>Item Variability (P (1-P))</u>	<u>Percent Variance Contributed</u>
1	50	.25	.38
2	80	.16	.24
3	20	.16	.24
4	10	.09	.14
5	100	<u>0</u> .66	0

Although we assigned a weight of only 1 to Item 1, it had more effect on ranking students, in this situation, than any other item analyzed. On the other hand, Item 5 with a weight of 3, actually had no effect or weight in ranking students.

The point of this illustration is that when you weight test items you may not obtain the results you desire. If a task element is important enough to weight, perhaps you should score the element on a go/no-go basis instead.

There is one other reason why we do not weight test items. It is an enormous administrative burden. We do not believe the results are worth the effort. Until we are able to agree on an overall task standard, why bother with elements within the task? After we discuss the problem of establishing standards, we would like your viewpoints on weighting test items.

In the Army we are concerned with how the soldier's on-the-job performance compares with an acceptable performance standard. On-the-job standards are usually expressed in terms of speed and/or accuracy. In the Army School System we are equally concerned with job standards, but are hampered by the following constraints:

- a. Nonexistence of job standards.
- b. Cost effectiveness.

The first constraint seems to cause the biggest problem. Often job

standards are not stated precisely enough for criterion testing and, therefore, require further analysis. For example, the job standard requiring military typists MOS (71B30) to type at a rate of 40 words per minute must be further defined. What will constitute an error? How many errors are allowable? Are erasures or pen and ink corrections acceptable? These questions must be answered before firm training standards are established. Since some errors may be permitted, successful performance of a task does not necessarily imply that perfection is required on each element of the task.

This problem is further compounded when subjective tasks such as briefings or staff studies are considered. A method has not yet been devised to break these tasks into their component elements and score them on a go/no-go basis. However, some standard must be established for training and testing. Normally this is accomplished by using the opinions of instructors, staff members and other qualified personnel.

When a standard is subjectively established, care must be taken to insure that the standard is based on the job requirement of a new person, not of an experienced incumbent.

On-the-job training and job experience are always factors in developing total job competency. Therefore, you must insure not only that your graduates meet the minimum standard expected in the field, but also that they do not exceed the job performance required at entry level. To exceed the expected standard must be considered overtraining. (This does not imply, however, that certain critical tasks will not require 100% performance; only that the analyst must be certain that tasks labeled critical are, in fact, essential for mission success or safety considerations.

In addition to the above, an absolute standard can rarely be established until the training system has been tried on several control groups. Since time and resources may be the overriding constraints, these control groups may be actual classes. The initial trial of a new instructional system may reveal many problem areas in the instructional design, test instruments, time allocations, etc. Therefore, it may be advisable to establish two standards initially:

a. Target Standard - student level of performance that represents what the entry field standard is under an optimum instructional system.

b. Interim Standard - the level of performance a new instructional system will support.

Unless adequate time is available to refine the instructional system, it may be necessary to rely on an interim standard to determine the initial pass/fail point. Except for reasons of chance, it can be assumed that the overall performance of a class must be a function of the quality of the instructional system. Therefore, until this system has been optimized, we could rely on norm-referenced procedures to establish any interim pass/fail points used.

After additional administrations of the test (usually between five and ten classes) this pass/fail point (interim standard) will become more stable and should equal the target pass/fail point. It should also become a reliable measure of an individual's achievement in a particular task, making more obvious those who actually did not profit from the instruction. After the quality of the instruction has started to level off, the target pass/fail point should be re-analyzed by qualified personnel with field experience to insure its relationship to actual field conditions. Later, the results of postgraduate and supervisor surveys can be used to confirm the original and subsequent estimates of the job standard. If the target pass/fail point is considered below the acceptable job standard, it may be necessary to increase the existing instructional resources devoted to the task. Conversely, if the performance standard is too high, action should be taken to reduce the training time spent on this task and apply it to a critical task or one considered deficient.

Once the training system has been optimized, the major concern of a training school must be to prepare students to perform a job-task at an entry level standard. If individual student performance is at or above the minimum prescribed standard, he passes; otherwise, he fails and should be considered for remedial instruction, recycling or elimination as shown by Inclosure 1. Since this system of grading is identical to that used on obstacle courses, it is commonly referred to as Multiple Hurdle Grading. The clearance of each hurdle often within a given time, constitutes a single training objective. Under the pure criterion testing concept, it does not matter whether he cleared the hurdle by 2 feet or 2 inches.

However, since it may be necessary to use overall class averages to determine class standing, promotion eligibility, commandant's list, etc., a modified system of multiple hurdle grading may be necessary. Under this system a student must successfully jump each hurdle before graduation from a course. Class standing is determined by his relative class average.

LISTED BELOW IS A SUMMARY OF THE DISCUSSION PERIOD THAT FOLLOWED THE ABOVE PRESENTATION.

1. Since many school-trained tasks cannot be compared to an absolute real-world standard, any pass/fail point established is only a best estimate.
2. As long as the requirement exists to rank students, pure criterion testing cannot be used. However, a mix of norm and criterion testing might be the best solution to the problem. Critical or vital tasks would be scored on a go/no-go basis and require a 100% standard. Noncritical tasks would be scored on a pass/fail basis similar to the method described above.
3. Considerable discussion was generated concerning the difficulty of a test item and whether the results of a test should have a "proper spread." It was pointed out that tests used in training situations are intended to monitor the quality of the training and pinpoint instructional problems. Therefore, the procedures commonly used in aptitude testing are not appropriate to use in courses designed under the systems engineering concept.
4. Several participants indicated that existing CONARC policy does not allow remedial instruction, retests, etc., to be accounted for as program of instruction (POI) time. Therefore, not only students but also instructors, must conduct this training after normal duty hours. It was suggested that this time, if accounted for properly, could be used as a reward to the successful student. Additionally, this change would no longer penalize the instructor, or the schools who utilize mandatory remedial instruction, if the current procedures were modified to account for this time in the POI.

MULTIPLE HURDLE GRADING

PASSING AREA

PASS

FAIL

IMMEDIATE RETEST (1)

REMEDIAL INSTRUCTION
THEN RETEST (2)

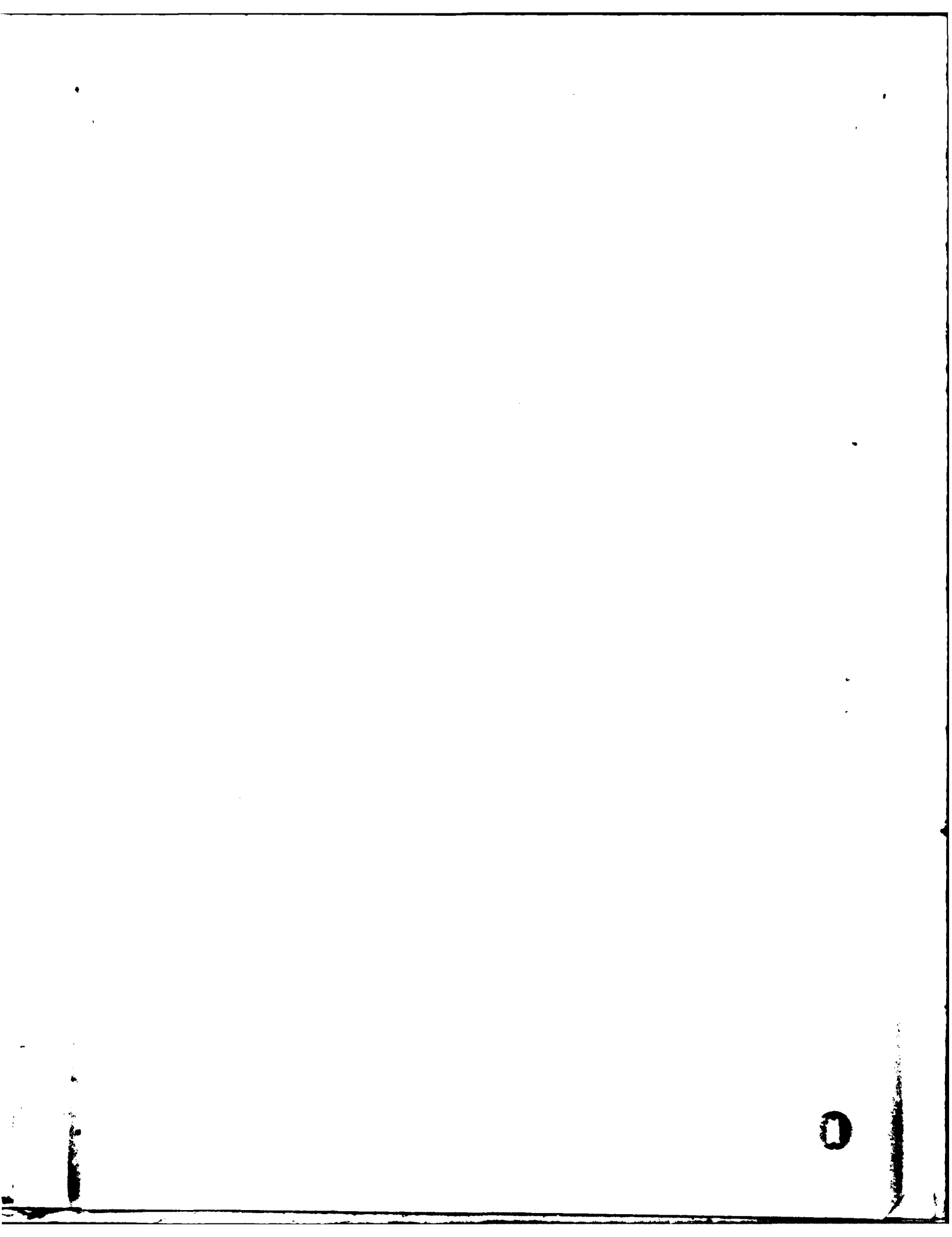
ELIMINATED FROM COURSE/ (3)
RECYCLE

Unless the task being tested is critical and requires 100% performance, each student who fails a task would be retested as shown above:

a. Students in area (1) would be provided a critique and immediately retested using an alternate version of the test instrument.

b. Students in area (2) would be provided a critique, remedial instruction and retested using an alternate version of the test instrument.

c. Students in area (3) would be provided a critique, and retrained as in paragraph b above, or recycled to another class, or eliminated from the course depending on the student's progress to date.



QUALITY CONTROL WORKSHOP
CRITERION - REFERENCED TESTING

By JAMES A. SQUIRES
US Army Air Defense School
Fort Bliss, Texas

W. James Popham credits Alfred Binet with launching what is now known as norm-referenced testing, but it is perhaps more significant that Popham emphasizes that development and use of norm-referenced tests can be credited to H. H. Goddard, a director of the Vineland, New Jersey Training School for the Feeble-Minded. I am not suggesting that norm-referenced tests should be relegated only to the feeble-minded; however, those involved in testing with sound minds should open those minds and consider the merits of criterion-referenced testing. Today, we are going to examine the merits of criterion-referenced testing, but, first, we must make note of norm-referenced testing and then enter it.

In his book, Educational Psychology, volume 1, Thorndike wrote, "... the report to the individual of his school marks was not the vice of the old system. Its vice was its relativity and indefiniteness -- the fact already described that a given mark did not mean any definite amount of knowledge, or power, or skill -- so that it was bound to be used for relative achievement only" The "old system" referred to by Thorndike in 1913 is still the system primarily used today in the military as well as in civilian schools. Though there is merit in the old system, the Armed Forces, with heavy emphasis on technical and vocational courses, realize little benefit from norm-referenced measurement.

At the US Army Air Defense School at Fort Bliss, Texas, some 56 courses of instruction are offered; yet 90 percent of these courses are concerned primarily with skill development. Surveys of the larger schools of the Navy, Army, Air Force, and Marines indicate that this high percentage of skill development courses is approximated throughout the schools of the Armed Forces. These schools must concern themselves with two primary postulates of training or educating. First, the student must arrive at the school with certain minimum skills and knowledges, or he must be taught these minimum skills and knowledges, and then he must absorb the instruction that is to prepare him for a post-graduation job. This means that, as a student, he must indicate his level of proficiency or knowledge not once, but twice. Once to qualify for the course of instruction, and once to qualify for the job. It is universally agreed among tests

and measurements personnel that norm-referenced tests do not indicate how proficient a student is with respect to the performance involved, but how proficient one student is relative to another student. Granted, most experts agree that criterion levels of proficiency or knowledge can be required of students who are norm-reference tested. However, this could be equated to setting the criteria for sprinters that they run the 100 yard dash in 9.5 seconds; after the race you would know who won and how many sprinters bested 9.5 seconds, but the individual time of each sprinter would be lost in the obscurity of relative standards.

So, since norm-referenced measurement and criterion-referenced measurement are both valid and accepted by testing authorities, what type measurement is best for your use? Alfred Garvin presented one simple test or criterion that can be used -- if at least one of the instructional objectives of a unit of instruction envisions a task that must subsequently be performed at a specific level of competence, in at least some situation, then criterion-referenced measurement is relevant. Stated negatively -- only if there is no criterion is criterion-referenced measurement irrelevant. However, today's instructional technology development and today's emphasis on skill development dictate that criterion-referenced measurement can be and must be used in most testing situations.

As is emphasized by this workshop and by others throughout the civilian and military communities, the name of the game today is systems engineering. Systems engineering of most courses precludes the use of norm-referenced measurement and emphasizes the necessity of criterion-referenced measurement. Systems engineering decrees that training objectives be established that are job oriented. This, in effect, establishes tasks which must be performed at a specific level of competence. The thousands of pages of Mager, Skinner, Garvin, Glaser, and others are testimonials to the job-oriented performance objective. Appendix A is an example of how systems engineering has necessitated the need for criterion-referenced measurement. Appendix A indicates that previously administered norm-referenced exams required that a student perform at a percentage level of 70 in order to complete this portion of the course and thereby qualify as a technician for this portion of his job. Neither the student, the instructor, test personnel, nor job supervisors could equate 210 points, 240 points, or 270 points to the student's ability to perform at a specific level. Two hundred and seventy points usually meant this student could be expected to perform better than the student with 210 points. However, even this statement was not necessarily true if the duties were isolated. For instance, a student with 270 points may have been deficient in performing checks and adjustments; whereas, a student with 210 points may have superior knowledge or skill in performing checks and adjustments, but may exhibit a weakness in other areas. I could inundate you with examples

such as this one from the systems engineering of one single course. This example indicates, perhaps tautologically, the validity of Garvin's original postulate -- if a task must be performed at a specific level of competence, then criterion-referenced measurement is relevant.

Systems engineering has given emphasis to a new or perhaps rebirth of a very old educational philosophy -- individualized or self-paced instruction. Prominent educators during the 35 years preceding 1970 proclaimed the recognition of individual differences as the prescription for better education while our public school machinery was attempting to inoculate all students with a virus of commonality in education. Fortunately, the pendulum is returning and the practical wisdom of adherents to individualized instruction is being emphasized once again. Modern instructional technology dictates the need for individualized, self-paced instruction. Individualized or self-paced instruction precludes the use of norm-referenced measurements since students and instructors must be concerned with absolute levels of proficiency, not relative levels compared to other students. A student has, ipso facto, become a class within himself. Individualized instruction follows Leslie Briggs' ideas that time and/or methods-media should be distributed, not scores.

Finally, criterion-referenced measurement has become the necessary instrument for testing because it is so well suited to the most prominent type of instruction today -- technical or vocational. As Briggs stated, technical training, with possibly a very few exceptions, is designed to train students to attain specified objectives satisfactorily. These specified objectives that are tested may be a domain of tasks, such as troubleshooting problems, but whether the technical instruction is isolated into a single task or domains of many tasks, they are tasks with a specified minimum acceptable criteria. At the Air Defense School, we teach approximately 56 courses of instruction, yet some 90 percent of these courses can be considered technical-vocational. Throughout these courses, it is necessary to test students to determine if they possess the minimum skills and knowledges to advance to other sections of the course. Wherever information is needed as to the adequacy of an individual's performance, criterion-referenced testing must be used.

Although it is not within the scope of this presentation to dwell upon the use of statistics in an overall exam program it is believed that some mention of these statistical items as applied to criterion-referenced testing must be made. Jackson, Livingston, Garvin, and others have encountered many problems involved in using exam statistics with criterion-referenced measurement. Livingston was optimistic when he stated, "With a few modifications, the classical theory of test reliability can be applied to criterion-referenced measures...."

However, he admits that neither he nor others have a presently proven substitute for the classical exam statistics. Most statistical manipulations are predicated on the use of norm-referenced testing which by its nature must be discriminatory, yielding a wide range of scores. Means, standard deviations, etc., are dependent on this variability. Scores such as percentile ranks, stanines, and grade equivalents lose the specificity inherent in criterion information. Variability, reliability, validity, and discrimination are not very important in terms of precisely what does a student know and how well is he prepared to do a specific job. Instead of being tied to the past with exam statistics that do little for education and training but do expend large sums of money for computers and statistical personnel, we should be following such guidelines as expounded by Robert Glaser. He has substituted "absolute measurements" for the "relative measurement" of the past. Absolute measurement is in step with the present and the future. This dictates that we measure a student's accomplishment in relation to an established standard. Each student must be compared with the singular or domain standard set specifically for him. Appendix C indicates how individual standards and absolute measurements can be used for skill level differentiation, setting of minimum standards, or student segregation. A new set of statistical data is needed for criterion-referenced measurement, and previous research into this area by Garvin, Bandura, Ebel, Cox, Glaser and others has indicated that educators and statisticians are most likely going to have to shed the blindfold of standard testing statistics before they can perceive a clear plan for criterion-referenced statistics. There is no valid reason for an educator to delay inauguration of needed criterion-referenced measurement while awaiting the birth of a reliable replacement.

To this point, I have been concerned with teasing you with the ideological, educational benefits of marrying a measurement system to a well-mated, modern instructional philosophy. Through the use of authoritative quotes, personal experience, and educational assimilation, it is easy to elicit many agreeing nods from your heads, but, heretofore, it has been ideological, and perhaps even visionary; however, there are many personal, specific benefits for each of you which can make "switching instead of fighting" very desirable. You are the managers, the supervisors, the innovators of instruction. You supervisors are faced daily with the task of pouring over pages of nebulous exam statistics, blindly juggling item analysis with item validity to arrive at a solution that will satisfy the student, mollify the instructor, and convince the manager. With the absolute interpretation of learning indicated by criterion-referenced measurement your dead-ends and detours will be eliminated. Your only concern is to teach the student to achieve the minimum standard or to indicate he cannot be taught the minimum standard.

You managers now have a time-equipment-cost breakdown on each student for each task and domain of tasks. This information should allow you to make more valid, authoritative decisions as to school training or on-the-job training; as to need for more hours of instruction or more instructors; and as to the worth of purchasing more equipment or teaching more hours.

You training innovators now have detailed information on each student accomplishment of each task. With this minute control you can vary one item at a time to determine the best method, the best media, the ideal time, and the ideal vehicle for maximum learning. You are now being given the opportunity to become true clinical psychologists for curriculum development.

If the goals are so rewarding, if the gains are so high, then why haven't we all clamored for criterion-referenced testing? Without a doubt the greatest stumbling block is the attitude of "rather fight than switch." This attitude too often prevails at all levels but it can be overcome with a concerted effort on the part of instructors, supervisors, educators, and managers. An unwavering course that leads us to criterion-referenced measurement must be set, and though the barriers may appear unyielding, it is the educator's job -- our job -- to finally make the attainment of knowledge more important than the attainment of a position within a group.

APPENDIX A

4B-F7/150-25x20 FIRE DISTRIBUTION SYSTEMS REPAIR

ANNEX A = 400 POINTS
ANNEX B = 350 POINTS
ANNEX C = 250 POINTS

TOTAL COURSE = 1000 POINTS

<u>STUDENT</u>	<u>ANNEX A</u>	<u>ANNEX B</u>	<u>ANNEX C</u>	<u>COURSE AVG</u>	<u>TOTAL POINTS</u>
1	70	80	60	71	710
2	60	90	90	78	780
3	60	60	100	70	700
4	75	80	80	78	780
5	100	65	65	79	790

APPENDIX B

ANNEX A.

TOTAL = 300 POINTS

PERFORM DAILY CHECKS

PERFORM ADJUSTMENTS

RECOGNIZE SYMPTOM OF MALFUNCTION

FAULT ISOLATION OF MALFUNCTION

REPAIR OF MALFUNCTION

MINIMUM ACCEPTABLE STANDARD = 210 POINTS

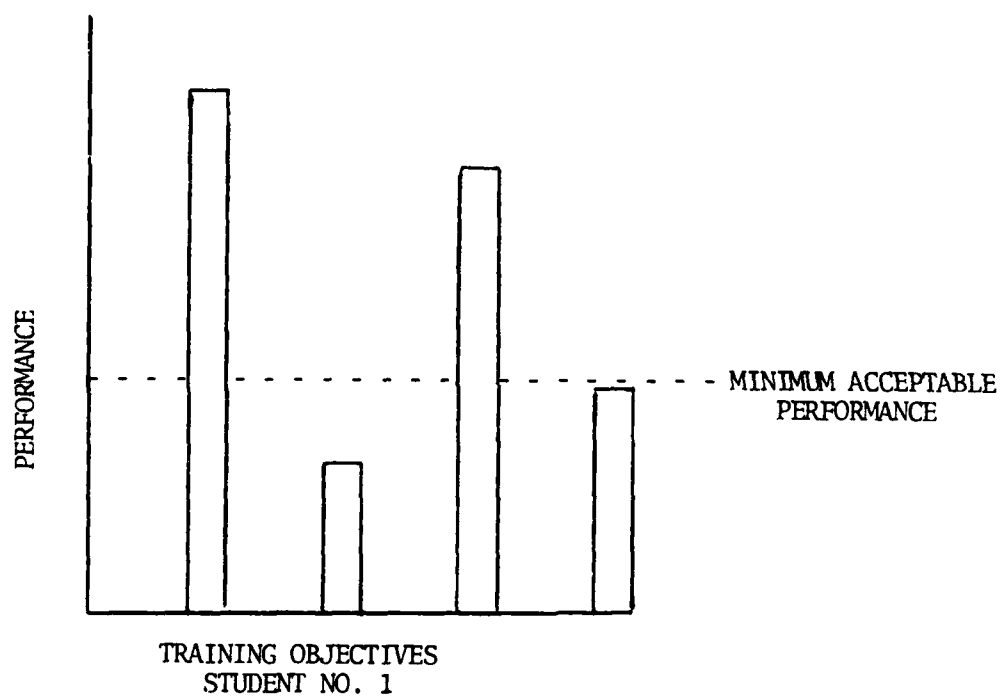
STUDENT A ACHIEVED 270 POINTS

STUDENT B ACHIEVED 240 POINTS

STUDENT C ACHIEVED 210 POINTS

STUDENT D ACHIEVED 200 POINTS

APPENDIX C



BIBLIOGRAPHY

- Block, James H., Criterion Referenced Measurements: Potential. School Review, Feb. 1971.
- Briggs, Leslie J., Handbook of Procedures for the Design of Instruction. American Institute of Research, 1970.
- Ebel, Robert L., Criterion Referenced Measurements: Limitations. School Review, Feb. 1971.
- Garvin, Alfred D., The Applicability of Criterion-Referenced Measurement by Content Area and Level. University of Cincinnati, 1970.
- Glaser, Robert and Nitko, Anthony J., Measurement in Learning and Instruction. University of Pittsburg, 1970.
- Jackson, Rex, Developing Criterion-Referenced Tests. Educational Testing Service, Princeton, N.J., June, 1970.
- Kriewall, Thomas E. and Hirsch, Edward., The Development and Interpretation of Criterion-Referenced Tests. American Educational Research Association, Feb. 1969.
- Lindvall, C. M. and Glaser, Robert., The Role of Evaluation in Individually Prescribed Instruction, University of Pittsburg, Feb. 1966.
- Livingston, Samuel A., The Reliability of Criterion-Referenced Measures. Johns-Hopkins University, July 1970.
- Myers, Charles T., A Covariance Approach to Item Analysis. Educational Testing Service, Princeton, N.J., Sep. 1970.
- Nitko, Anthony J., Criterion-Referenced Testing in the Content of Instruction. University of Pittsburg, Oct. 1970.
- Popham, W. James, Editor, Criterion-Referenced Measurement. Educational Technology Publications, Englewood Cliffs, N.J., 1971.
- Popham, W. James and Husek, T. R., Implications of Criterion-Referenced Measurement. Journal of Educational Measurement, Spring, 1969.
- 31.2

QUALITY CONTROL WORKSHOP

GO/NO-GO GRADING

By EARL CARR
US Army Southeastern Signal School
Fort Gordon, Georgia

(Slide 1 On)

Ladies and Gentlement - The purpose of this presentation is to provide you with a brief overview of the Southeastern Signal School's Go/No-Go Grading Program.

(Slide 1 Off)

Go/No-Go testing and grading evolved from the need to assure that school graduates can perform satisfactorily upon job entry. In the past, most tests in this school were multiple-choice written tests which tested subject matter knowledge rather than performance skills. There were some actual performance tests, but they were not always related to the full scope of job requirements. The development of behavioral objectives, related to actual job performance led to a testing and grading policy which was designed to directly reflect job entry performance capability. The intent was to graduate only those students who could show evidence, on job sample tests, that they could satisfactorily perform upon job entry.

These basic principles have been refined and reinforced in the Continental Army Command Regulations on

(Slide 2 On)

(1) Systems engineering of training, and

(Slide 3 On)

(2) Army Schools curriculum policies.

(Slide 3 Off)

In keeping with these policies, the Southeastern Signal School began refining its training and testing programs. Currently over 95% of the tests in our school's enlisted MOS-producing courses are

job-sample performance tests. In most cases the actions required of a student in a test situation are the same actions required upon job entry. For example, if the job task is to align radio XYZ, test performance requires the student to actually align the radio. Particular emphasis is placed on employing realistic job-related performance standards. In the case of aligning the radio, test standards, or driteria, specify that minimum acceptable performance requires that the radio meet field operational standards as specified in the technical manual. A student cannot pass by completing 70 or 80 percent of the procedural steps required in alignment. The radio must be ready for operational use.

This situation that I have described, typifies criterion-referenced or go/no-go testing at the Southeastern Signal School. If a student meets performance test criteria, he is allowed to "go" or proceed to the next training objective. If he cannot perform, he is not allowed to go forward. If he cannot perform he will receive remedial training, and he must satisfy all course objectives before he can graduate.

In a test program of this nature, it is obvious that the conventional normative-referenced or "curved" grading method is not appropriate. The grading system must be "go/no-go" also. To this end the CONARC regulation previously cited on Army Schools Curriculum Policies states, and I quote, "Go/no-go grading is encouraged where appropriate. Local programs will be established to phase out grading systems in which the comparison of students to their contemporaries is the significant factor and/or those which use curved grading." End quote. In theory, go/no-go grading means no grades will be given. But without grades, how do we meet the student ranking requirements for identification of honor graduates and promotion eligibles? We must have an administrative record of academic performance for these purposes. But certainly we must not modify academic performance requirements just for the sake of grades or ranking.

Student ranking at this school is accomplished by using two methods: one method for students in group-paced courses; another method for students in self-paced courses.

(Slide 4 On)

Let me first discuss student ranking in group-paced courses and begin by pointing out that "grades" are still used. In this case it is the manner in which the grade is derived that becomes significant. The grade must be kept in the go/no-go context.

(Slide 4 Off)

In actual practice grades are derived by converting the minimum acceptable performance criteria to a value or grade of 70. For performance exceeding the minimum acceptable, additional points are given to provide grades up to 100. The 70 to 100 could be other numbers which represent minimum and maximum performance. We used 70 and 100 to facilitate record keeping in the transition from curve to go/no-go testing. Now to be more specific, a test standard (criteria) may require that the student align one of two radios within a certain time frame. If one is aligned to meet operational requirements -- not to a certain percentage of the procedural steps involved in the alignment, but to meet operational requirements--the grade would be 70; if both are aligned in the same time frame the grade would be 100. It should be noted that the additional points for maximum performance here, and in most cases, are based on repetitions of the required performance rather than on a greater degree of perfection with regard to the action requirement. This approach provides proof of student accomplishment of the objective, while requiring less than 100 percent efficiency. Less than 100 percent efficiency is considered acceptable for job entry, however, inability to perform to the objective requirement is not acceptable. After all test grades are obtained, final course averages for ranking purposes are conventionally determined by obtaining the arithmetic grade average.

(Slide 5 On)

Student ranking in self-paced courses is accomplished in a slightly different manner. Student Test performance still relates directly to criterion-referenced standards, as previously described, but in this case no grades are given for each test. Student ranking is accomplished by computing each student's course completion time and number of test failures. The student completing a course in the least time with the fewest test failures will be the highest ranking student.

(Slide 5 Off)

The actual computation is done in two steps. In the first step, a Progression Index is derived for each student.

(Slide 6 On)

This is done by dividing the student's actual training time in hours excluding absenteeism, by the allotted training time.

(Slide 7 On)

As the first example shows, if a student completes a 500-hour course in 500 hours, his Progression Index, or PI, is 1.00.

(Slide 8 On)

Example two shows the results of course completion in 400 hours and

(Slide 9 On)

Example 3, a thirty percent overage in training time.

(Slide 9 Off)

The inclusion of the Progression Index in the ranking procedure obviously reflects the rate at which the student performed in the course. Because of this the Progression Index is used for other administrative purposes in addition to ranking. For example, certain supplemental training and counseling actions are taken when a Progression Index reaches 1.1 and 1.2 respectively. Experience in two courses indicate that students who exceed a Progression Index of 1.3 usually do not complete the course. Because of this an in-course committee assessment of the student's progress and potential is made when the Progression Index reaches a 1.3 and the student may be relieved from the course. (Additional information is available to you on use of the Progression Index for administrative purposes in the Southeastern Signal School's presentation on Management and Control of Self-Paced instruction.)

Following the development of the Progression Index, the second step of the ranking procedure is performed. This involves substituting the actual Progression Index and the number of test failures into the formula shown on the slide.

(Slide 10 On)

Pause - In the formula, the numerical values are constant and the PI and F represent the Progression Index and the number of test failures respectively. The 100 simply puts the formula on a base of 100, which I will explain further in a moment. The first 2 is used to produce a value within the parenthesis which increases as the progression rate decreases. In other words, the faster student receives a lower PI, and therefore the resultant value within the parenthesis is higher, which produces a higher ranking score. The last 2 is used to increase the weight of a test failure in the ranking score.

(Slide 11 On)

As the example shows, the formula is designed to produce a ranking score of 100 for a student who completes a course in the allotted training time and has no test failures. In this case the Progression Index would be one and the number of test failures zero.
- Pause -.

(Slide 11 Off)

As previously stated, the student performance required on every test in the course was based on criterion-referenced standards that duplicate field entry requirements. The student can do the job required of him, and in training he did it in the prescribed time with no failures; thus a ranking score of 100.

To provide another example, let's assume for a moment that a student completed a 500-hour course in 400 hours and had one test failure.

(Slide 12 On)

You would first compute his Progression Index and find it to be .80. Then, second, you would substitute this Progression Index of .80 and the one test failure into the academic ranking score formula. After computation you would find the score to be 118. You will note that this student finished the course in less than the allotted time and that his score was greater than 100. In other words, the shorter the training time, the higher the ranking score.

(Slide 12 Off)

(Slide 13 On)

In another example, a student who requires 600 hours to complete a 500-hour course would have a PI of 1.2. If that student also had three test failures throughout the course, then the three failures would be inserted into the academic ranking score formula along with the PI of 1.2. After computation, the ranking score would be 74. In this example, excessive training time and test failures reduced the ranking score below 100. - Pause -.

(Slide 13 Off)

The test failure factor in the formula serves primarily to reduce the ranking score of a student who had test failures but who may have finished the course in the same time as another student who had no failures.

In summary, the School's testing and grading program, in both group-paced and self-paced courses, provides for job-sample performance based on criterion-referenced or go/no-go standards. As you have seen emphasis is placed on doing the job, and the rewards are for doing the job efficiently. The Southeastern Signal School believes that with go/no-go testing the student who enters his assignment will be better prepared to do the job than ever before. This concludes my presentation.

SLIDES FOR PRESENTATION
"GO/NO-GO GRADING
AT THE
US ARMY SOUTHEASTERN SIGNAL SCHOOL"

Slide 1 -

GO/NO-GO GRADING
AT USASESS

Slide 2 -

CONARC REGULATION
CONARC REG 350-100-1
SYSTEMS ENGINEERING OF TRAINING
COURSE DESIGN, 1 FEB 1968

Slide 3 -

CONARC REGULATION
CONARC REG 350-1, ANNEX Q
ARMY SCHOOLS CURRICULUM
ADMINISTRATION AND TRAINING
POLICIES, 15 DEC 1969

Slide 4 -

STUDENT RANKING
IN GROUP-PACED COURSES

Slide 5 -

STUDENT RANKING
IN SELF-PACED COURSES

Slide 6 -

$$\begin{array}{l} \text{PROGRESSION} \\ \text{INDEX} \\ \text{(PI)} \end{array} = \frac{\begin{array}{l} \text{ACTUAL} \\ \text{TRAINING} \\ \text{TIME (HRS)} \end{array}}{\begin{array}{l} \text{ALLOTTED} \\ \text{TRAINING} \\ \text{TIME (HRS)} \end{array}}$$

Slide 7 -

$$PI = \frac{\text{ACTUAL TRAINING TIME (HOURS)}}{\text{ALLOTTED TRAINING TIME (HOURS)}}$$

$$\text{EXAMPLE 1: } PI = \frac{500}{500} = 1.00$$

Slide 8 -

$$PI = \frac{\text{ACTUAL TRAINING TIME (HOURS)}}{\text{ALLOTTED TRAINING TIME (HOURS)}}$$

$$\text{EXAMPLE 1: } PI = \frac{500}{500} = 1.00$$

$$\text{EXAMPLE 2: } PI = \frac{400}{500} = .80$$

Slide 9 -

$$PI = \frac{\text{ACTUAL TRAINING TIME (HOURS)}}{\text{ALLOTTED TRAINING TIME (HOURS)}}$$

$$\text{EXAMPLE 1: } PI = \frac{500}{500} = 1.00$$

$$\text{EXAMPLE 2: } PI = \frac{400}{500} = .80$$

$$\text{EXAMPLE 3: } PI = \frac{650}{500} = 1.30$$

Slide 10 -

ACADEMIC RANKING = $100 (2-PI) - 2F$
SCORE (ARS)

WHERE: PI = PROGRESSION INDEX

F = NUMBER OF TEST FAILURES

Slide 11 -

ARS = $100 (2-PI) - 2F$

EXAMPLE: ARS = $100 (2-PI) - 2F$
 = $100 (2-1) - (2 \times 0)$
 = $100 (1) - 0$
 = 100

Slide 12 -

RANKING COMPUTATION

1. $PI = \frac{\text{ACTUAL TRAINING TIME (HOURS)}}{\text{ALLOTTED TRAINING TIME (HOURS)}} = \frac{400}{500} = .80$

2. ARS = $100 (2-PI) - 2F$
 = $100 (2-.8) - (2 \times 1)$
 = $100 (1.2) - 2$
 = $120 - 2$
 = 118

Slide 13

RANKING COMPUTATION

1. $PI = \frac{\text{ACTUAL TRAINING TIME (HOURS)}}{\text{ALLOTTED TRAINING TIME (HOURS)}} = \frac{600}{500} = 1.20$

2. ARS = $100 (2-PI) - 2F$
 = $100 (2-1.2) - (2 \times 3)$
 = $100 (.8) - 6$
 = $80 - 6$
 = 74

QUALITY CONTROL WORKSHOP

PROMOTION OF ENLISTED PERSONNEL AND DESIGNATION OF HONOR GRADUATES

By MAJ WILLIAM V. GREEN
Chief, Curriculum & Evaluation Division, USAAGS

DA Message 905026, dated 15 April 1969, governs the promotion of students in Army Service Schools. The current policy of promoting the upper 50 percent of a graduating class, provided they are otherwise eligible, is followed by the AG School. However, on their end-of course critique sheets, graduates have complained about the inequity inherent in the existing promotion policy. This inequity can be best illustrated by examining two classes of the Automatic Data Processing Machine Operator Specialist Course (MOS 74D20) taught by the US Army AG School.

Class	Total Graduated	Total Promoted	Avg Class Score	Score of	
				Indiv Promoted Highest	Lowest
#11	14	7	89.4	96.7	88.5
#12	14	7	95.1	99.4	96.3

Consequently any individual in Class #12 who scored below 96.3 was ineligible for promotion, while an individual in Class #11 who scored 88.5 or above was promoted.

This same kind of inequity exists at the AG School in all six enlisted courses where this promotion policy applies: Personnel Management Specialist, Postal, Computer Programmer, ADP Machine Operator Specialist, Computer Operator and Stenography.

To illustrate a solution to this inequity, the highest and lowest scores of those promoted in five classes of one course were plotted on graphs (Chart 1). Since the upper 50 percent of each class is eligible to be promoted, the lowest score promoted would be the median score for that class.

An average of these scores would provide an equitable promotion goal. Only those students scoring above this goal would be promoted. Using five classes of the Computer Programmer Course as an example, the same number of students would have been promoted using the median average as were promoted using the "upper 50 percent of class" rule. For the

classes surveyed in the other courses, more students would have been promoted using the median average. Nevertheless, over a longer period of time the number promoted using the median average should equal, or nearly equal, the number promoted under the 50 percent rule. Furthermore, those promoted would have performed at the same, or near the same, level of proficiency and the great variance in overall class scores of those promoted would be eliminated.

For each course, the first median average, or goal, would be established using a full year's class median average. Under the moving average principle, as each class graduates, its median score would replace the median score of the oldest class and a new median average would be computed. By maintaining a computer listing tape for each course, this update and computation could be easily accomplished through automatic data processing.

CONCLUSION: a. Inequity in the current promotion system does exist. An individual in one class of a course with a 96.2 average was not promoted while an individual in a different class of the same course with an 88.5 average was promoted.

b. An average of class median scores would provide a more equitable promotion cutoff point. Only those students scoring above that promotion score would be promoted.

Using the average median score method, students promoted would have performed at the same, or near the same, level of proficiency and the overall class averages of those promoted would not vary drastically from class to class.

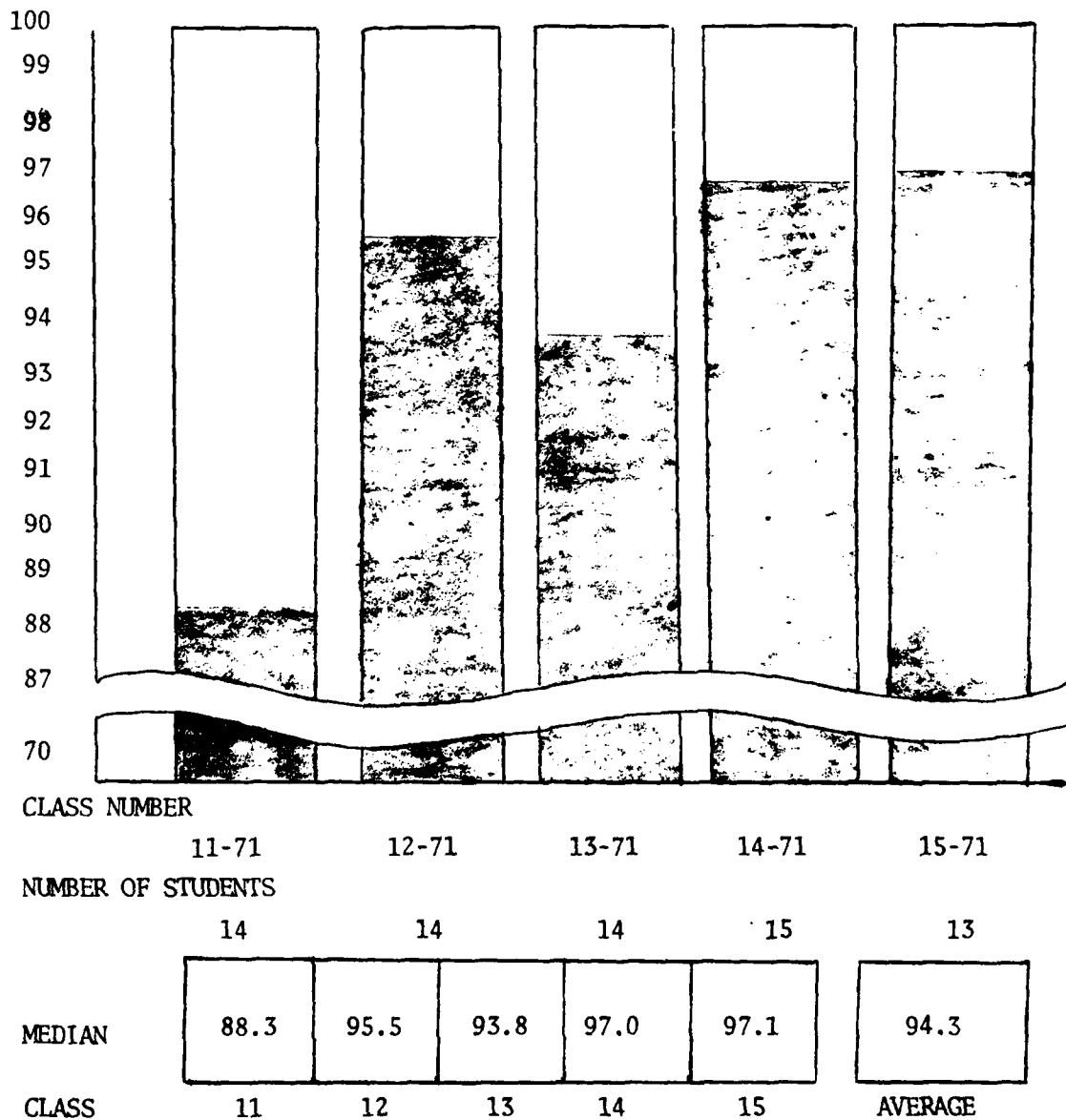
LISTED BELOW IS A SUMMARY OF THE DISCUSSION PERIOD THAT FOLLOWED THE ABOVE PRESENTATION.

a. There was general agreement that an absolute goal, rather than relative class standing, should be used to determine the eligibility for promotion of enlisted personnel and determination of academic achievement (i.e., honor graduates, commandants list, etc).

b. All participants agreed that the elimination of promotion as a reward for academic achievement would have an adverse effect on morale.

c. Participants could not agree, however, on a system that could replace the existing promotion rule. It was recommended that CONARC be informed of the problem, with recommendations to design a more equitable system.

(74D20)



Using this system the median-average (94.3) would be the promotion goal for Class 1-72. Only those students scoring above this goal would be promoted.

QUALITY CONTROL WORKSHOP

EXTERNAL QUALITY CONTROL

By MAJ WILLIAM V. GREEN
Chief, Curriculum & Evaluation Division, USAAGS

Prior to this workshop we conducted a survey of all the CONARC schools to determine what types of Quality Control information were used. The chart below shows the consolidation of this survey.

RESULTS OF QUALITY CONTROL SURVEY

	<u>% OF SCHOOL USING</u>	<u>RANK/ORDER</u>
TEST RESULTS	90	1
AUDITORS'/EVALUATORS' REPORTS	60	2
POST GRADUATE QUEST	85	3
STUDENT COMMENT SHEETS	95	4
MOD BANK REPORTS	45	5
MOS TEST RESULTS	40	6

Note that test results are used by 90% of the schools and ranked "one" as far as their utility or use.

Let's briefly discuss the uses of the other types of feedback listed.

Postgraduate Questionnaires - We attempt to use postgraduate surveys to update our original task inventory. In addition to asking the graduate if he performs a task, we also ask him to rank the task in relationship to all the tasks he performs. However, when processed (or averaged), all task values tend to drift toward the mean of 3.5. This last item results in a suspicious statistic.

We have also tried asking other questions, such as: How often do you perform the task? How long does it take you to perform the task? However, we have found little use or faith in any of the data processed, except for the percentage of incumbents performing a task.

In addition to postgraduate surveys, we also ask the graduates' supervisor to complete a questionnaire. We ask the supervisors such questions as: How much additional on-the-job training do you provide for each task listed? How do you think the task should be trained: OJT, formal school training, etc? Do you require incumbents to perform the task? How would you rank our recent graduate in comparison to all your personnel?

One problem we have found with supervisor questionnaires is the "halo effect" or "leniency error". For example, most of the supervisors all agree 90% of our graduates are in the upper 20%. This problem can be resolved to some extent by interviewing the supervisors in person; however, cost prohibits us from conducting all of our surveys by interview.

The major problem we have with questionnaires is the low return rate. For example, we are averaging only a 40% return rate for our officer and a 30% return rate for our enlisted graduates. I would like to know if anyone here has found a way to increase their return rates?

STUDENT COMMENT SHEETS - I'm sure all of us use comment sheets to attempt to evaluate our training, but most of us find them of little use for the following reasons:

- (1) Students tend to be too generous.
- (2) Opinions do not give direct evidence of their ability to perform.
- (3) Students are not familiar with on-the-job requirements.

We have found student comment sheets useful, to some extent, to indicate instructor performance. However, we still feel test results are the best indicator of instructor proficiency. Therefore, we feel the student is in no position to judge the quality of the training he received.

MOS TESTS - MOS test results can be used by some of the CONARC schools for particular courses. However, since MOS tests are administered to a person only after he has been in the service for 23 months, we cannot use them for many of our high density or entry level training courses. Additionally, many MOS tests are still "knowledge oriented" and do not reflect the person's true capability to perform on-the-job. Shown below is an example of this problem:

CORRELATION BETWEEN TROUBLESHOOTING (TS) AND
CORRECTIVE ACTION (CA) ON MOS TESTING AND
WORK-SAMPLE (WS)

	<u>TS-WS</u>	<u>CA-WS</u>
MOS 63C30	.33	.12
MOS 63C40	.46	-.02

As indicated, this recent HumRRO Study found little correlation between job performance and MOS test results. The more performance oriented our MOS tests are made, the more predictive they will be of job performance.

AUDITORS'/EVALUATORS' REPORTS - Many schools indicated they use auditors and course evaluator experts to furnish feedback. However, due to our lack of qualified subject area auditors, we have found them of little use in evaluating the objectives of systems engineering. We do, however, use this information to monitor our instructor techniques.

MILITARY OCCUPATIONAL DATA BANK - The major problem we have had with the current MOD Report is the infrequency of surveys. For these reports to be useful, they must be prepared more often than the Central Data Bank allows. For example, before MODB II can be implemented, the majority, if not all, of the service school courses will be beyond the job analysis phase of systems engineering. Therefore, MODB II would be beneficial to service schools for quality control purposes only. Additionally, it is estimated that because of the numerous MOS, each MOS will be surveyed once every four or five years. During four or five years many changes could occur in the MOS structure, equipment and doctrine. Therefore, the survey results would be of negligible value for use by the service schools. Thus, schools will be forced to continue processing postgraduate and pre-systems engineering surveys, and will continue to duplicate the function of a centralized data bank.

LISTED BELOW IS A SUMMARY OF THE DISCUSSION PERIOD THAT FOLLOWED THE ABOVE PRESENTATION.

There was considerable discussion that indicated schools should use all types of feedback to control the quality of their training. It was pointed out that no one quality control source is a cure-all.